Science

2008
Visualization Challenge

AAAS

COVER

Boulders and roots, or spores and hyphae? Parallel microscopic and macroscopic worlds merge in this detail from one of the winners of this year's *Science*/NSF International Science & Engineering Visualization Challenge. See the special section beginning on page 1767.

Image: Colleen Champ, with micrographs by Dennis Kunkel

DEPARTMENTS

EDITORIAL

NEWS OF THE WEEK

1758

NEWS FOCUS

LETTERS

BOOKS *ET AL.*

1779

EDUCATION FORUM

PERSPECTIVES

Science

SCIENCE EXPRESS

www.scienceexpress.org

REVIEW

BREVIA

1801

RESEARCH ARTICLES

REPORTS

Science

1755 & 1828

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

Printed on
30% post-consumer
recycled paper.
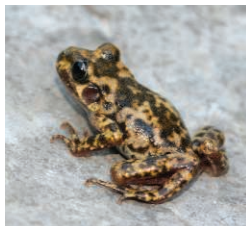
Toad troubles.

SCIENCE**NOW**
www.sciencenow.org
HIGHLIGHTS FROM OUR DAILY NEWS COVERAGE

Conservation Efforts May Have Backfired for Spanish Toad
A potentially deadly fungus was accidentally introduced as part of a breeding program for an endangered amphibian.

Wasps Make Peace With Past Enemies
The insects steer clear of foes they have fought in the past.

Fat Molecule Fights Weight Gain
Compound prevents mice from storing unhealthy fat.

Intracellular colocalization of Grb10 and IGF-1 receptors.

SCIENCE **SIGNALING**
www.sciencesignaling.org
THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

RESEARCH ARTICLE: Nedd4 Controls Animal Growth by Regulating IGF-1 Signaling
X. R. Cao, N. L. Lill, N. Boase, P. P. Shi, D. R. Croucher, H. Shan, J. Qu, E. M. Sweezer, T. Place, P. A. Kirby, R. J. Daly, S. Kumar, B. Yang
Nedd4 acts through Grb10 to enhance insulin-like growth factor signaling and control animal growth.

PERSPECTIVE: Caspase-2—Vestigial Remnant or Master Regulator?
C. M. Troy and E. M. Ribe
Both mitochondrial-dependent and -independent cell death pathways are mediated by caspase-2.

PODCAST
E. M. Adler, N. R. Gough, A. M. VanHook
Bacteria secrete factors that regulate genes that contribute to virulence.

Dave Jensen tools up his own career.

SCIENCE **CAREERS**
www.sciencecareers.org/career_development
FREE CAREER RESOURCES FOR SCIENTISTS

Tooling Up: My Career Dissected
D. Jensen
Dave Jensen highlights some of his own career mistakes.

From Watching 'The Expert' to Being an Expert
L. Cahoon
Melanie Lee attributes her success partly to a childhood TV show.

Union Aftershocks in California
C. Rey
Postdocs and others in the University of California system adjust to life with a union.

The *Science* Careers Web Log
Science *Careers Staff*
Here's where to find information from around the Web on careers in the sciences.

SCIENCE**PODCAST**
www.sciencemag.org/multimedia/podcast
FREE WEEKLY SHOW

Download the 26 September *Science* Podcast to hear about Earth's potentially oldest rocks, science policy and the U.S. presidential election, this year's International Science & Engineering Visualization Challenge, and more.

Separate individual or institutional subscriptions to these products may be required for full-text access.
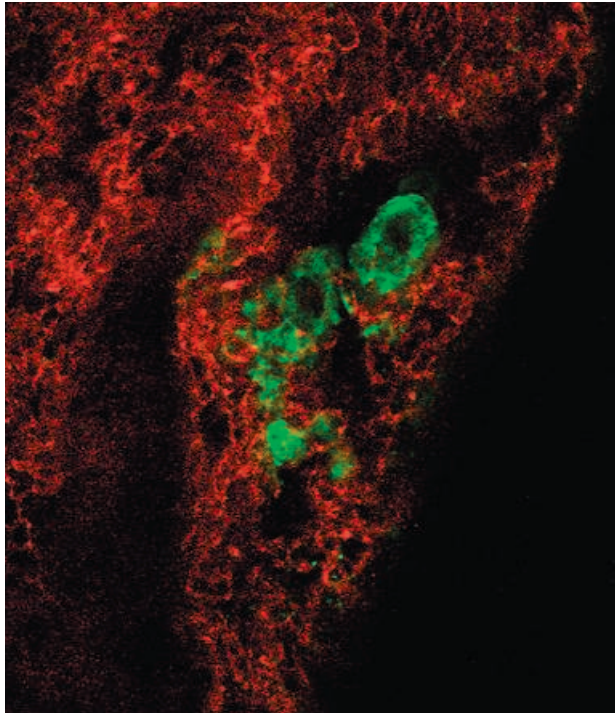
<< Rethinking Cancer Metastasis

Most human cancer deaths are caused by metastasis, in which cancer cells spread from the primary tumor to new sites in the body. Because metastatic cells must successfully negotiate a series of complex steps, including survival in the bloodstream and establishment in a foreign tissue environment, metastasis has been viewed as a late event in cancer progression **Podsypanina et al.** (p. 1841, published online 28 August; see the Perspective by **Klein**) suggest that the metastatic process may begin earlier than previously thought. Normal mouse mammary cells were genetically manipulated to allow the timing of oncogene expression to be experimentally controlled and injected into the bloodstream of mice. Surprisingly, in the absence of oncogene expression, normal mammary cells were capable of traveling to and surviving in the lungs for up to 16 weeks, although they did not initiate aggressive growth until after oncogene activation. Thus, metastases might arise from disseminated normal (premalignant) cells that remain clinically silent until genetic changes render them malignant.

DNA Templates for Nanomachinery

The precise and complementary base pair matching in DNA has increasingly led to its use as a building or templating material in the assembly of nanoscale objects like particles or wires, or for the decoration of particles and wires with metals or other molecules. **Aldaye et al.** (p. 1795) review recent developments in the use of DNA as a precise positional tool for complex material assembly. Developments have moved from simple one-dimensional templating to two and three dimensions, with scope for dynamically changing the shape or size of an object, or the fabrication of nanomachines.

Cancer Genomes: From Chaos Comes Order?

Identification of the genes altered in cancer cells is critical for understanding how the disease arises and for designing more effective diagnostic tests and therapies (see 5 September news story by **Kaiser**). **Parsons et al.** (p. 1807, published online 4 September) and **Jones et al.** (p. 1801, published online 4 September) catalog the numerous genomic alterations that help turn normal cells into two of the deadliest human cancers: glioblastoma multiforme (the most common type of brain cancer) and pancreatic cancer. Although for each cancer type, the specific genomic alterations varied from tumor to tumor, the altered genes affected a limited number of cellular signaling pathways and

regulatory processes, suggesting that these are the pathways that go awry and lead to the disease. Of particular interest in the glioblastoma study was the discovery of recurrent mutations in the active site of isocitrate dehydrogenase 1, encoded by the *IDH1* gene. In this small study, *IDH1* mutations were more prevalent in glioblastomas from younger patients and in "secondary" glioblastomas, and they were associated with a better prognosis.

Martian Dynamo

One surprise from recent spacecraft observations of Mars is that its crust in the southern hemisphere is strongly magnetized, but not so in the northern hemisphere. This pattern seems similar to the major crustal difference on Mars in that the northern hemisphere is relatively smooth, at a much lower elevation, and younger. Mars now lacks an active dynamo. **Stanley et al.** (p. 1822; see the news story by **Kerr** and the Perspective by **Langlais and Hagay**) show through numerical models that if the heat flow were lower across the core-mantle boundary in the northern hemisphere, as might be expected from any mechanism producing the crustal dichotomy, the resulting geomagnetic field might not be a dipole but be concentrated just in the south. Such a dynamo would also affect Mars' atmospheric evolution because only part of the planet would be strongly shielded from the solar wind.
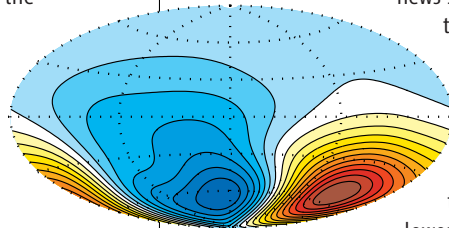
Working Together to Get the Job Done

Bob tries to make a call to Alice but finds that the line is too noisy. Picking up his second phone (he's a very busy builder), he finds that line is also too noisy and so gives up trying to contact her. With two bad lines, Bob wouldn't be able to make that phone call, at least using the classical communication channels of his provider. Had he had access to quantum communication channels, **Smith and Yard** (p. 1812, published online 21 August; see the Perspective by **Oppenheim**) show theoretically that the situation is quite different. Two quantum channels, each with zero capacity to transmit information independently, will allow information to be carried across them when used together. Not only of theoretical interest, this counterintuitive result may be of practical use in the design of quantum communication networks.

Dissecting a Disordered Material

Graphite oxide was first prepared almost 150 years ago, but the functionalization of the graphite is not uniform, which has hampered efforts to characterize it. This material is now of interest as a precursor for the formation of graphene, which has potentially useful electronic properties. **Cai et al.** (p. 1815) have now prepared graphite oxide from graphite with varying degrees of ^{13}C-labeling (up to almost 100%). The labeled product allowed much higher resolution in solid-state nuclear magnetic resonance studies and excluded some of the potential models for the chemical bonding network of this material.

Sodium's Nonlinear Response

The influence of solvent rearrangements on chemical reactions in solution is often modeled using the linear response approximation, which essentially dictates that all starting configurations that equilibrate to a given final state do so with the same dynamics. **Bragg** *et al.* (p. 1817; see the Perspective by **Stratt**) show that the approximation comes up short for the formation of neutral sodium-electron ion pairs in tetrahydrofuran. Equilibration is twice as fast when the reaction proceeds by reduction of a Na^+ precursor than when Na^- is oxidized. The breakdown can be attributed to the large size differences between the cation, anion, and neutral, which substantially alter the extent of necessary solvent cavity rearrangements in each case.

Modeling Ocean Circulation

Hydrothermal systems along ocean ridges help control the chemistry of the oceans and alter and hydrate the upper oceanic crust; this, in turn, returns water to the Earth's mantle at subduction zones. Hydrothermal systems also foster deep ocean ecosystems. Observations seem to indicate that although ocean ridges are broadly linear, outflows are spaced out along them**Comou** *et al.* (p. 1825) have developed a three-dimensional numerical model of this flow to help reveal the dynamics. Their model shows that optimizing heat transfer causes the flows to self-organize into narrow pipe-like upflows, spaced about 500 m apart, fed by zones of warm downflow that recirculate up to a quarter of the heat.

Cystic Fibrosis Remodeled

Cystic fibrosis (CF) is caused by mutational disruption of*CFTR*, a gene encoding an ion channel required for chloride- and bicarbonate-mediated fluid secretion in epithelia and for salt absorption in

many organs. Two decades of intense research on CFTR has not yet translated into new clinical therapies, in part because mice—the traditional animal model for human disease research—do not develop the full spectrum of pathologies seen in human CF. To address this problem, **Rogers** *et al.* (p. 1837) have inactivated the *CFTR* gene in pigs, an animal that shares many anatomical and physiological features of humans. Newborn pigs lacking CFTR developed many of the gastrointestinal pathologies seen in infants with CF, including intestinal obstruction and abnormalities of the pancreas, liver, and gallbladder, and their nasal epithelia showed defects in chloride transport. These results, while still preliminary, suggest that the pig model may be a valuable tool for testing new therapies for CF.
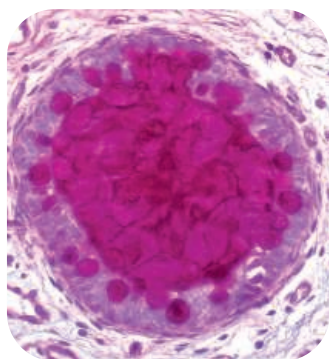
From the Minds of Babes

Human babies between 8 months and a year of age cannot perform certain cognitive tasks. In one of these, called the A-not-B error, an object is hidden under a container and the infant repeatedly reaches for it. Then the experimenter hides the object under a different container, in full view of the infant, but the baby still looks under the first container to find it. **Topál** *et al.* (p. 1831) propose a new explanation for this error, suggesting that the socially intense "teaching" interaction that usually accompanies the repeated hiding of the object under the first container ensures strong association of the object with that location. When the object is hidden without any communication between the experimenter and the infant, the baby's error rate is reduced. Previous explanations for the phenomenon suggested that it was due to the immaturity of the infant's executive motor control or his or her limited cognitive capacities.

The Agony of Defeat

Auctioneers take advantage of human nature to increase the sale prices of items. But are they banking on the successful bidder's enjoyment of winning, or are they instead relying on the bidder's aversion to losing? Two sides of the same coin, one might say, but **Delgado** *et al.* (p. 1849; see the Perspective by **Maskin**) argue that it is the latter that drives the phenomenon known as overbidding. When participating in an auction, brain areas sensitive to loss became active. When the authors modified the ground rules of the auction so as to emphasize the potential for loss, without altering the basic possibility of winning, the tendency to overbid was magnified.

John Edward Porter is a former U.S. congressman who chaired the Appropriations subcommittee that funds all federal health programs, including NIH. He is chair of Research!America and chaired the U.S. National Academy of Sciences committee that has just published a report advising presidential candidates on science and technology appointments.*

If All You Do Is Vote …

ELECTIONS HAVE A WAY OF SORTING THINGS OUT. ALREADY, THE MOST FASCINATING U.S. presidential and congressional election process of my life has sorted out some things that we can celebrate. We know that the country's next president won't favor teaching intelligent design in our schools and will respect scientific integrity and evidence-based research. But we still don't know whether he will truly put science at the table—that is, whether research will be high on the president's priority list and reflected strongly in his budgets, speeches, and policies.

So what can U.S. scientists do to substantially increase the probability that we will have elected officials who will make research a very high priority? I'm talking about much more than voting on Election Day, paying dues to a professional society, or making a contribution to a voluntary health association. And here's why.

For the past 7 years, the United States has had a presidential administration where science has had little place at the table. We have had a president opposed to embryonic stem cell research and in favor of teaching intelligent design. We have had an administration that at times has suppressed, rewritten, ignored, or abused scientific research. At a time when scientific opportunity has never been greater, we have had five straight years of inadequate increases for U.S. research agencies, which for some like the National Institutes of Health (NIH) means decreases after inflation.

All of this has been devastating for the scientific community; has undermined the future of our economy, which depends on innovation; and has slowed progress toward better health and greater longevity for people around the world. So if you are a U.S. scientist, what should you do now?

First, help identify candidates for the next president's science appointments. They range across a variety of agencies and departments, and the U.S. National Academies have listed those viewed as most important.* Urge your distinguished colleagues to serve our nation in this way, and help the scientific community to support them.

Second, in choosing candidates who are running for Congress or even state office (often, state officials will later run for federal office), volunteer to advise those candidates on science matters and issues. They'll love it! Offer to serve on their science advisory committee. If they don't have one, tell them you'll create one. Chair it yourself and recruit suitable colleagues. Once your candidate has won the election, offer to continue in your role as a science adviser. Wouldn't it be wonderful if all candidates had science advisers or science advisory committees? They will, if individual scientists step up to the plate.

Third, school yourself on the candidates and their positions on science issues. Visit science voter education resources, like YourCandidatesYourHealth.org, which asks all federal candidates to answer questions about their positions on science and health. If your candidates have not responded, call their campaigns and ask them to do so. You have a right to know where they stand.

Fourth, encourage debates about science among those who seek public office. Go to their debates and raise science questions. Sign onto ScienceDebate2008.com, which urges the presidential candidates to have a debate dedicated to science issues. Even though this won't happen now, support for this initiative will send a message to the media and the candidates that science is important to the electorate and that the questioners should include science in the debates.

And last but not least, next time run for office yourself! It's disheartening to see so many public officials with little knowledge of science. Bill Foster, a physicist, recently won the House seat of former Speaker Dennis Hastert. You can do it, too.

Your country needs you. If all you do is vote, you're definitely not doing enough. Get off your chair, do something outside your comfort zone, and make a difference for science. All of us must be creative about what we can do to make a difference for the things we believe in. Now is the time.

– **John Edward Porter**

*Science and Technology for America's Progress: Ensuring the Best Presidential Appointments in the New Administration, www.nap.edu/catalog.php?record_id=12481.
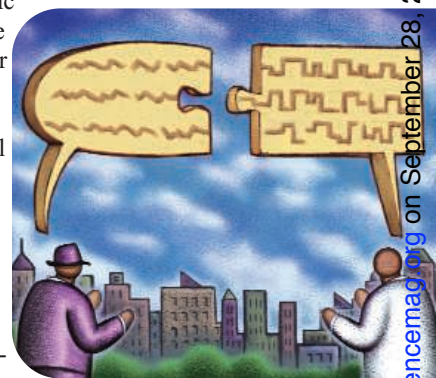
CELL BIOLOGY

A New Way in

Many cellular stimuli induce signaling cascades that terminate with a protein entering the nucleus to activate transcription of target genes. Most of these proteins contain a conserved stretch of amino acids known as a nuclear localization signal (NLS), which binds to the nuclear import factor importin alpha, and the complex translocates into the nucleus through the nuclear pores. However, the absence of an NLS in some signaling proteins suggested that they access the nucleus via alternative mechanisms. Now, Chuderland et al. find a new signal in the extracellular signal–related kinase 2 (ERK-2). A three–amino acid domain is phosphorylated upon stimulation, allowing the protein to bind to a different nuclear import factor, importin7, and enter the nucleus. A similar domain was found in other cytonuclear shuttling proteins, and the same phosphorylation-dependent mechanism was shown to occur for nuclear accumulation of SMAD3 and MEK1. Thus, this domain acts as a general nuclear translocation signal and represents a new mechanism whereby proteins can enter the nucleus. — HP*

Mol. Cell **31**, 10.1016/j.molcel.2008.08.007 (2008).

CHEMISTRY

More Surface, More Reactivity

To gain a better understanding of palladium's reactivity as a hydrogenation catalyst, many model studies that use well-defined single-crystal surfaces have focused on what should be the simplest substrate, ethylene. Although this reaction is facile for reactant pressures near ambient, at very low pressures (ultrahigh-vacuum conditions), the reactivity on close-packed surfaces is low (yields of ethane <1%), and not much greater on supported nanoparticles (<5%). This difference is attributed to a lack of surface hydrogen caused by absorption into the bulk. Dohnálek et al. have prepared model catalysts through ballistic deposition of Pd atoms at cryogenic conditions (22 K) and glancing angles such that one-quarter of the atoms are surface exposed. The as-prepared nanoporous films showed much higher reactivities (50%), which decreased when the films were densified by reaction cycles that went to room temperature (with ethane desorbing by 250 K) or after annealing to higher temperatures. The authors note that although surface roughening treatments can also create a large number of active sites, the low fraction of bulk atoms in the nanoporous films limits removal of hydrogen from the surface and boosts overall reaction rates. — PDS
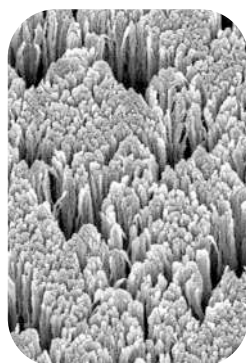
J. Phys. Chem. C **112**, 10.1021/jp803880x (2008).

CELL BIOLOGY

Growing Through a Wall

In bacteria, the cell wall must be firm enough to define cell shape and allow a high internal osmotic pressure, while at the same time sufficiently dynamic to allow cell growth and division. Hayhurst et al. provide insight into how the cell wall in the rod-shaped organism, *Bacillus subtilis*, is structurally organized to achieve these functions. The main structural component of the cell wall is peptidoglycan, comprising glycan strands cross-linked by peptides. Atomic force microscopy (AFM) on purified glycan revealed individual strands up to 5 µm long (5000 disaccharides). Fluorescence microscopy in whole cells showed that *B. subtilis* displays very few terminal *N*-acetyl glucosamine (GlcNAc) residues and that internal peptidoglycan-associated GlcNAc residues exhibit a pattern suggestive of a helical structure. AFM imaging of *B. subtilis* peptidoglycan sacculi revealed little indication of structural features on the outer surface, probably because surface layers are hydrolyzed during cell wall turnover. However, the inner surface exhibited 50-nm-wide cables running across the short axis of the cell with cross striations consistent with a helical structure. The authors suggest that during biosynthesis, glycan strands are polymerized and cross-linked and then coiled to form the inner-surface cables. New helices are likely inserted into the cell wall by being cross-linked between two existing cables, while the external surface is cleaved to allow cell growth. — VV

Proc. Natl. Acad. Sci. U.S.A. **105**, 14600 (2008).

DEVELOPMENT

Protected by a Maelstrom

Germ-line cells could be considered the most precious in the body, because they are the only cells to contribute directly to the next generation. Hence, special mechanisms should be in place to protect them from damaging agents such as transposable elements. Cells in many species silence these elements by using small noncoding RNAs. The RNA interference factors localize to perinuclear structures called nuage in germ cells. Soper et al. focus on a murine homolog of *Drosophila* maelstrom (*mael*), a gene that functions in the production of interfering RNAs, repression of transposable elements, and specifi-

cation of the *Drosophila* oocyte axis. Similarly, the murine *Mael* gene is localized stage-specifically to the nuage structures in male germ cells. Eliminating *Mael* from mice resulted in defective meiosis due to abnormal chromosome synapsis and massive DNA damage. A mechanism for meiotic failure is demonstrated through *Mael*'s function in transcriptional repression of transposable elements via a DNA methylation mechanism. — BAP

Dev. Cell **15**, 285 (2008).

CELL BIOLOGY

NE-ER Shape Shifting

Mitosis in metazoans involves the wholesale disruption of normal cellular architecture to allow for successful partitioning of cellular components to each daughter cell. During most of the cell cycle, the nucleus is surrounded by a double-membraned nuclear envelope (NE) that is contiguous with the endoplasmic reticulum (ER), an intracellular labyrinth of interconnected tubules and sheets. Anderson and Hetzer have examined the processes involved in the dramatic rearrangements of the NE and ER at the end of mitosis. The NE is disassembled at the beginning of mitosis and, after the partitioning of chromosomes, must be reassembled to form two daughter cells complete with their own NE-enclosed nuclei. By quantifying images produced using time lapse microscopy, the authors were able to observe the recruitment of ER tubules to chromatin, which went on, within ~12 min, to produce membrane-enclosed daughter nuclei capable of performing nuclear import. Increasing the expression of ER tubule–promoting proteins interfered with the formation of new nuclei, whereas reducing their expression sped up the process, which may suggest that it is the transition of ER from tubules to sheets that limits NE assembly and nuclear expansion. Thus, ER architectural proteins play a key role in nuclear reconstruction and NE assembly after mitosis. — SMH

J. Cell Biol. **182**, 911 (2008).

ER proteins (green) gathering about chromatin (red).

CLIMATE SCIENCE

A Hurricane History

One problem in assessing whether recent climate change has significantly influenced either the strength or frequency of hurricanes and tropical storms is that in general, these factors have been measured systematically only recently. Thus, establishing a reliable baseline to compare with present trends has been difficult. In the Lesser Antilles—one of the first areas settled heavily in the New World, and a focus of early trade—British ships' logs, newspaper accounts, official colonial correspondence, and other sources provide a variety of data over the past 300 years or so. Chenoweth and Divine used these sources to derive a historical record of hurricanes, tropical storms, and tropical depressions in this region, which is along the main track of storms that eventually develop and hit the United States and Mexico. The authors identified 550 tropical storms and hurricanes passing through these islands, about half of which were not previously detected, including in more recent records. Overall, there seems to be no discernable trend in activity since 1690, though the period from 1968 to 1977 had notably few storms. — BH

Geochem. Geophys. Geosyst. **9**, 10.1029/2008GC002066 (2008).

BIOCHEMISTRY

SH2 Uninhibited

Src-homology 2 (SH2) domains of cytoplasmic tyrosine kinases have an important and well-defined role in keeping such kinases in an autoinhibited conformation. Filippakopoulos *et al.* studied the human cytoplasmic tyrosine kinase Fes, which lacks this autoinhibitory interaction, and uncovered molecular details of how SH2 domains can alternatively act to enhance activity. The authors solved crystal structures of a portion of Fes containing the SH2 domain and kinase domain, with and without phosphorylation of the kinase activation segment. This fragment was bound in complexes with a substrate peptide and an ATP-mimetic kinase inhibitor. Mutagenesis experiments confirmed that the visualized interaction of the SH2 domain with the kinase domain was necessary to stabilize the active conformation of the enzyme. Analysis of synthetic substrates with or without phosphorylated SH2 domain–binding sites also showed the importance of the SH2 domain in substrate recruitment. Extending the analysis to the pro-oncogenic tyrosine kinase c-Abl showed that a similar mechanism occurs in other members of the cytoplasmic tyrosine kinase family. The authors point out that such coupling of substrate recognition to kinase activation may contribute to selectivity of such kinases so that they act only on the appropriate substrates in vivo. — LBR

Cell **134**, 793 (2008).

The Galápagos' Lost and Found

Of the 15 species of the iconic Galápagos giant tortoise, four have already gone the way of the dodo. Charles Darwin wrote about the relentless culling of one now-extinct species on Floreana Island, *Geochelone elephantopus*, for food and lighting oil. In a second chapter of this story, a team from Yale University reported in this week's *Proceedings of the National Academy of Sciences* that they have found 13 descendents of this species alive and well on neighboring Isabela Island.

"It's strange that the human activities responsible for depleting this population have allowed it to survive elsewhere," says senior author Gisella Caccone. She theorizes that whalers left the tortoises on Isabela's highest summit, Volcan Wolf, as a larder to feed them on a return trip.

By comparing the DNA from 93 Volcan Wolf tortoises with that of museum specimens, the team discovered that 13 had some Floreana lineage. "It's extraordinary to find descendents of a species when it already has gone extinct," says George Amato, director of conservation genetics at the American Museum of Natural History in New York City. Caccone says it may even be possible to resurrect the species in captivity by selectively breeding out the Isabela genes.

True to Stereotype

Behind every stereotype lies a kernel of truth, the saying goes. A recent study of regional personality differences in the United States seems to agree. Jason Rentfrow, a psychologist at University of Cambridge, U.K., invited Internet users across the United States to take a survey devised to assess psychology's "Big Five" personality dimensions: openness, conscientiousness, extraversion, agreeableness, and neurotic tendencies, such as anxiety.

More than 619,000 people responded. Breaking the responses down by state, Rentfrow found that personality traits clustered by region:

"Neurotic" NYC cabbie?

Northeasterners scored highest on the neuroticism and openness scales but were not particularly conscientious or agreeable. Denizens of the Midwest and the South had the highest conscientiousness and agreeableness ratings. For extraversion, the Midwest, the South, and the Great Plains states ranked highest, Rentfrow reported in a paper published online this month in *Perspectives on Psychological Science*. Although participants were slightly younger than the general population, their racial and gender breakdown mirrored that of the country, Rentfrow says.

The study is the latest in a growing field that uses the Internet to study aggregate personality traits of entire populations. Robert McCrae, a psychologist at the National Institute on Aging in Bethesda, Maryland, who studies personality differences between countries, calls the study "a very ambitious attempt to get at a new way of analyzing personality data."

WoW! NSF Funded *What*?

Bloggers have been heaping scorn on the U.S. National Science Foundation (NSF) for awarding $100,000 to computer scientist Bonnie Nardi of the University of California, Irvine, to study the popular role-playing computer game World of Warcraft (WoW).

Users can freely make add-ons to the basic WoW software, and Nardi wants to know why American users edit the software more often than their Chinese counterparts do. But WoW players say they already know the answer: As one wrote in a post to the Web site GamePolitics, "More Americans play WOW on computers they own. More Chinese players play at internet cafes and computer centers."

But NSF is sticking up for its grant. "While we have previously supported research on highly formalized open-source software development," says NSF program officer William Bainbridge, "this may be the first study we have supported on how software is developed in the nearly complete absence of formal organization."

Probing Near-Death

Books have been written about "near-death" experiences—visions of tunnels and figures of light or accounts of hovering over the operating table watching doctors bang on their chests—occurring when a patient's heart and brain have stopped functioning. Now, a British physician is spearheading a large-scale project aimed at finding out what's going on.

The study of Awareness During Resuscitation, sponsored by the University of Southampton, U.K., was announced this month at a United Nations symposium on consciousness by project leader Sam Parnia, a resident at New York–Presbyterian Medical Center. Parnia has recruited 25 hospitals, mostly in the United States and the United Kingdom, to monitor as many as 1500 people during cardiac arrest who then survive to tell about it. "About 10% of such people report some kind of cognitive process" while "dead" for a few seconds to more than an hour, Parnia says.

Psychiatrist C. Bruce Greyson of the University of Virginia, Charlottesville, says emergency rooms and intensive-care units will measure oxygen flow to patients' brains and will test their blood for proteins released when brain cells die. Researchers will also ascertain whether patients accurately describe things from their out-of-body experiences that they could not have seen.

What if the phenomenon proves real? "I think that shows that the current understanding of brain and mind"—that to have such experiences you need "a coherent neural network involving a good portion of the cortex"—is "inadequate," Greyson says.

NEWS>>

THIS WEEK

More overruns for
Mars lander

1754

Q&A with Britain's
science adviser

1756

SCIENCE EDUCATION

Misjudged Talk Opens Creationist Rift at Royal Society

A talk titled "Should Creationism Be a Part of the Science Curriculum?" was bound to attract attention at the annual meeting of the British Association for the Advancement of Science (BA) earlier this month. But last week, it cost the speaker, Michael Reiss, his job as director of education at the Royal Society, Britain's academy of science.

Within hours of his 11 September talk, news items appeared on the Internet claiming that Reiss had urged science educators to teach creationism, although many attending the speech said that he had clearly not made such a call. His comments—or perhaps more accurately the spin placed on them by headline writers, newspaper columnists, and editorialists—ignited a firestorm. Several prominent scientists, including a trio of Nobel laureates, called for his resignation. The Royal Society hastily put out a statement defending Reiss but 4 days later issued another statement announcing his resignation and leaving the clear impression he had been forced out.

Although some critical of Reiss applauded his sudden exit, others saved their harsh words for the organization he left. "This has damaged the Royal Society, the way they handled it," says Derek Bell, head of the Association for Science Education in the United Kingdom.

On paper, Reiss, who remains a professor at the University of London's Institute of Education, seemed the perfect speaker at the science festival in Liverpool. In addition to having a doctorate in science education, he's an ordained minister in the Church of England and coedited the book *Teaching About Scientific Origins: Taking Account of Creationism.* In his 11 September talk, Reiss noted that teachers are bound to encounter pupils with creationist views. If these are brought up in class, he argued, simply dismissing them as not appropriate to a science lesson will only alienate those pupils. Instead, Reiss advocates taking the opportunity to explain the difference between the creationist viewpoint, which, he emphasizes, has no evidence to support it, and evolution, which, he says, has a lot. A teacher's answers to such questioning from creationist pupils "can be used to illustrate a number of aspects of how science works," Reiss says in the online text of his talk (www1.the-ba.net/bafos/press/showtalk2.asp?TalkID=301).

Efforts by creationists or believers in intelligent design to tamper with school science

Wise words? Reiss's comments on dealing with creationist pupils were jumped on by U.K. newspapers.
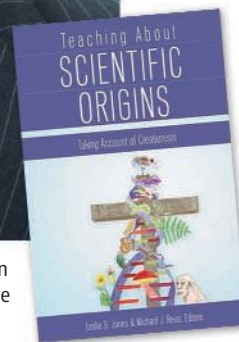
curricula are viewed in Europe as largely an American phenomenon, one reflected in the public acceptance of the theory of evolution. Only about 40% of adults in the United States accept the idea of Darwinism, compared with about 70% in the United Kingdom and many other European nations (*Science*, 11 August 2006, p. 765). Yet throughout Europe, groups promoting creationist views are emerging from Protestant, Catholic, and Islamic communities in different countries.

In the United Kingdom, government pol-

icy throughout this decade has been to create more so-called faith schools, high schools funded predominantly by government but managed by religious organizations. This has increased concern among science educators about the spread of creationism into curricula. There was also widespread condemnation in 2006 of a pressure group called Truth in Science that sent creationist teaching materials to every U.K. high school. In part because of this, in 2007 the government published guidance for schools on how to address creationism and intelligent design, drafted by Reiss and others.

The day after his speech, Reiss was greeted by headlines such as "Call for creationism in the classroom," in the *Financial Times*, and "Children should be taught about creationism in school," in the *Daily Mail.* Highlighting Reiss's description of creationism as a "worldview" that should be respected, many of the stories suggested he equated it with the theory of evolution. That same day, the Royal Society issued a statement reaffirming its position that creationism should not be taught as science, saying that Reiss's views had been "misrepresented" and offering a clarification from him.

That move failed to quell the storm. Several fellows of the Royal Society stated publicly that it wasn't appropriate for Reiss to hold such an influential position given his religious affiliation. "I do not see how he could continue," says Nobelist Harry Kroto of Florida State University in Tallahassee. On 13 September, another society fellow, Nobelist Richard Roberts of New England Biolabs in Ipswich, Massachusetts, sent a letter to the Royal Society, cosigned by Kroto and John Sulston of the University of Manchester in the U.K., calling for Reiss to step down.

Reiss has declined to comment since his talk, but on 16 September the society issued a new statement, saying that Reiss's comments "were open to misinterpretation. While it was not his intention, this has led to damage to the Society's reputation." The society and Reiss, the statement continued, had agreed that "in the best interests of the Society, he will step down immediately."

Scientists and experts in science education have leapt to Reiss's defense. "There's an awful

CREDIT: INSTITUTE OF EDUCATION

FOCUS

Time for a human
proteome project?

1758

KEEPING
AMERICA
COMPETITIVE

Science and the
2008 campaign

1762
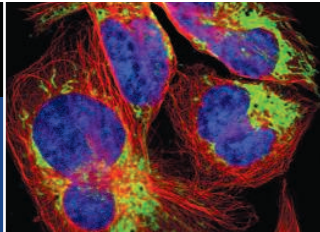
lot of support for Michael Reiss, as a person and his views," says Bell. Fertility expert and society fellow Robert Winston of Imperial College London issued a statement critical of the Royal Society: "This is not a good day for the reputation of science or scientists. This individual was arguing that we should engage with and address public misconceptions about science—something that the Royal Society should applaud." Even evolutionary biologist and noted creationism critic Richard Dawkins defended Reiss on his Web site, call-

ing efforts to remove him "a little too close to a witch-hunt."

Others, however, welcomed Reiss's departure as the best way for the Royal Society to make clear its position on creationism. "The only reason to mention creationism in schools is to enable teachers to demonstrate why the idea is scientific nonsense," says Christopher Higgins, vice-chancellor of Durham University in the U.K.

Royal Society President Martin Rees said in an e-mail to *Science* that "the Royal Soci-

ety should be secular, but not anti-religious." But Phil Willis, a member of the U.K. Parliament who chairs a committee that oversees British science, acknowledges that there is a "very stark division" between those in the society who reject religion completely and those who urge coexistence or are themselves religious. Although Willis has great respect for Reiss, he was "caught making injudicious comments," Willis says. "It's a real tragedy, [but] it's inevitable that they parted company."

–DANIEL CLERY

PARTICLE PHYSICS

After Spectacular Start, the LHC Injures Itself

When physicists first sent particles racing through the world's biggest atom smasher on 10 September, the Large Hadron Collider (LHC) at the European particle physics laboratory, CERN, near Geneva, Switzerland, the gargantuan machine purred like a kitten. But only 9 days later, the LHC proved it can also be a temperamental tiger, damaging itself so severely that it will be out of action until next spring.

It all seemed so easy earlier this month when after just a few hours researchers had beams of protons whizzing through both of the 27-kilometer-long, $5.5 billion machine's countercirculating rings. That smooth start raised hopes that the LHC would start colliding particles as early as this week. But last Friday, some of the 1232 main superconducting dipole magnets, which keep the beams on their circular trajectories, abruptly overheated in an event known as a "quench." The incident ruptured the plumbing that carries liquid helium through the magnets to chill them to 2 kelvin—2 degrees above absolute zero.

The quench highlights the LHC's ability to injure itself, says Reinhard Bacher of the DESY particle physics lab in Hamburg, Germany. Compared with earlier colliders, "the LHC is operating in all aspects much more at the critical edge," he says. The breakdown raises the question of whether a key protection system worked properly.

A superconducting magnet is

essentially a coil of superconducting wire that generates a magnetic field when current flows through it. If kept extremely cold, the wire carries huge currents without resistance; a quench occurs when part of it overheats and acts like an ordinary wire. The hot bit serves as an electric heater that can trigger a runaway reaction, toasting the rest of the magnet and converting the energy in its field to heat.

Such an event can start if, for example, protons stray out the beam pipe and into the magnet material. The 19 September quench occurred a different way, however. Within the LHC, the 15-meter-long magnets are connected so that current from one magnet flows into the next. With no beam, researchers were ramping up the current in a chain of 154 dipole magnets when the superconducting connection between two magnets apparently overheated and melted, says CERN spokesperson James Gillies. That breach can also cause an uncontrolled quench. The loss of current causes a magnet's field to quickly ebb. However, the magnet will produce a voltage surge

to counteract the waning of the field. The reaction can then heat the magnet as a whole.

The field of an LHC dipole contains a whopping 8.6 megajoules of energy, enough to melt 42 kilograms of copper or to cook the magnet in a fraction of a second. Researchers designed a quench-protection system to quickly shunt current out of an afflicted magnet and into large steel blocks. Ironically, it also heats the entire magnet to spread out the toll from the quench. During last Friday's mishap, the quench-protection circuits fired as expected, Gillies says.

Nevertheless, the broken helium line suggests that at least one $900,000 magnet was ruined and will need to be replaced. "I cannot tell if any of the safety systems failed," says CERN's Rüdiger Schmidt, who leads the machine-protection team. "It is absolutely too early to tell."

To make repairs, researchers will have to warm up an entire octant of the LHC and then cool it back down. Bacher says that DESY researchers experienced similar delays when they commissioned their HERA collider, which smashed electrons into protons from 1992 to 2007. "We had three or four of these warm-up-and-cool-down cycles," he says. One such cycle will take at least 2 months for the LHC. That will run into a planned shutdown for the winter, when the cost of power climbs, so experimenters will have to wait until next spring to collect data.

–ADRIAN CHO
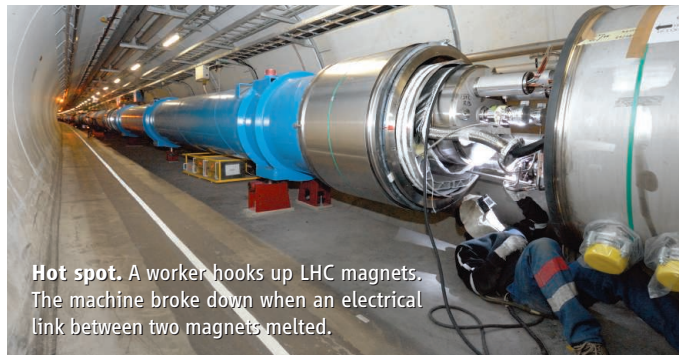
Hot spot. A worker hooks up LHC magnets. The machine broke down when an electrical link between two magnets melted.

CREDIT: CERN
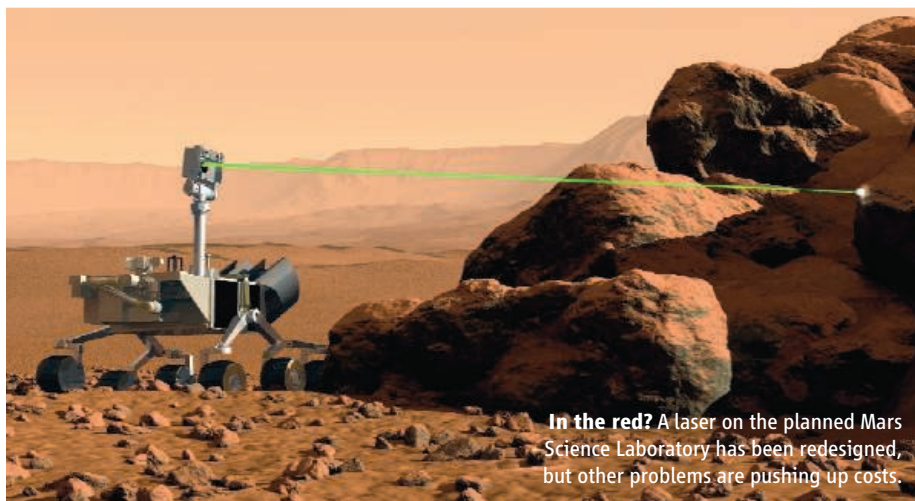
In the red? A laser on the planned Mars Science Laboratory has been redesigned, but other problems are pushing up costs.

SPACE SCIENCE

Rising Costs Could Delay NASA's Next Mission to Mars and Future Launches

Faced with a dramatically higher price tag, NASA managers will decide next month whether to postpone the launch of a sophisticated Mars rover for 2 years. Such a delay in the Mars Science Laboratory (MSL) would mark a significant setback to the Mars research program, which has sent a new spacecraft to the planet every other year for a decade. Planetary scientists also worry that pushing back the mission could have a ripple effect, delaying and even canceling future missions.

The science laboratory, currently slated for launch in the fall of 2009, is four times heavier than the current rovers trundling across the planet's surface. It features a plethora of advanced tools and instruments designed to analyze rocks, soil, and atmosphere. But that complexity has led to technical troubles and higher costs. When proposed in 2004, the lab was expected to cost $1.2 billion. By this summer, that price tag had climbed to $1.9 billion, and last week NASA space science chief Edward Weiler warned that "there is another overrun coming." Another NASA official put the latest increase at approximately $300 million.

Engineers at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California, which is responsible for MSL, are now working overtime to prepare the spacecraft for environmental testing and launch. Weiler and NASA Administrator Michael Griffin will pore over those results and the latest cost estimates when they meet in mid-October to determine whether to delay the mission. In addition to

worrying about the unbudgeted overtime, Weiler is concerned that engineers may be rushing their inspection of the rover's complicated systems. "The alternative could be that we get a crater on Mars," the science chief adds ominously, evoking previously failed missions to the Red Planet.

"Postponing MSL is a real possibility, and an unfortunate one," says Brown University planetary scientist Jack Mustard, who also chairs NASA's Mars science advisory panel. "A 2-year delay could increase the cost of the mission significantly—and that would come out of the Mars budget." He also fears that this could slow momentum for Mars exploration and jeopardize plans for a 2016 rover and a sample-return mission.

MSL's cost and technical woes began in earnest last year, when NASA considered jettisoning two instruments, ChemCam and the Mars Descent Imager. ChemCam, an instrument designed by French and American researchers that would use a laser to vaporize martian rock and dust for spectrographic analysis and take detailed photographs, proved more expensive than anticipated. In a compromise reached last fall, NASA provided extra money to complete a simplified version of the instrument. The Mars Descent Imager, a camera designed to provide a view from just above the surface of the rover's landing site, was canceled but revived when engineers found ways to control costs by making relatively

minor changes.

But those changes didn't stanch the financial bleeding. The latest technical problems affecting the MSL budget include the tardy delivery of hardware used in the sample acquisition and handling portion of the laboratory. NASA Planetary Science Division Director Jim Green said in June that the total overrun for MSL in 2008 and 2009 was $190 million. Most of that money—some $115 million—will come from other Mars-related projects. JPL spokesperson Guy Webster referred MSL questions to NASA headquarters.

A new $300 million overrun, says a NASA official familiar with MSL, could force the agency to cancel the $485 million 2013 Scout mission announced just last week to probe the planet's atmosphere (*Science*, 19 September, p. 1621) or the 2016 Mars mission. "Rest assured the Mars program has to pay for this," the official added.

Mustard, who last week chaired a Mars advisory panel session in California, says there is growing anxiety in the community about the implications of an MSL delay on an exploration program that began in the late 1990s. "If you delay it until 2011, then you might lose the 2016 mission," he says. The 2016 Mars effort now under consideration likely would be a smaller rover that could include some sample-gathering technology designed to test systems for an eventual sample-return mission from Mars to Earth. The projected $1.4 billion cost of such a rover would fall between MSL and the current Spirit and Opportunity rovers now on the surface.

NASA's former space science chief, S. Alan Stern, who resigned in protest this spring after a disagreement with Griffin over how to deal with cost overruns (*Science*, 4 April, p. 31), last year proposed a Mars sample-return mission to arrive at the end of the next decade. But static budgets, spacecraft overruns, and the need to conduct other missions make that increasingly unrealistic, say agency managers and academic researchers. Weiler notes that a sample-return mission would cost many billions of dollars and that NASA is planning first to launch a mission to either Jupiter or Saturn late in the next decade. And although scientists are intrigued by the idea of a sample-return mission, they see it slipping into the more distant future.

"Plans for a sample return were smoke and mirrors," says Mustard. "It's a good idea—but where's the money?" **–ANDREW LAWLER**
With reporting by Richard A. Kerr.

"Postponing MSL is a real possibility, and an unfortunate one."
—JACK MUSTARD, BROWN UNIVERSITY

GEOCHEMISTRY

Geologists Find Vestige of Early Earth—Maybe World's Oldest Rock

Really old stuff is rare on Earth. The planet's brand of violent geology has just been too dynamic to preserve much from its earliest days. Formed 4.567 billion years ago, Earth has yielded 4.3-billion-year-old mineral grains and 4.0-billion-year-old rocks that hint at how a ball of primordial debris evolved into a crusted-over, largely ocean-covered abode of life.

So geologists keep searching the oldest, most brutally battered terrains for more traces of earliest Earth. On page 1828, a group reports the discovery of rock in northern Quebec on Hudson Bay that records the existence of the earliest crust. The Canadian rock may also be the oldest known rock by 300 million years.

Given how beaten up the oldest rocks are, geologists often fall back on atomic-scale records preserved in the isotopic and elemental composition of the rocks. Geologist Jonathan O'Neil of McGill University in Montreal, geochemist Richard Carlson of the Carnegie Institution of Washington's Department of Terrestrial Magnetism in Washington, D.C., and colleagues analyzed isotopes of the elements samarium and neodymium from the Nuvvuagittuq greenstone belt of northern Quebec. These isotopes can be used to trace geologic processes because some isotopes are stable and don't change no matter how many eons pass, whereas some steadily decay radioactively into other, more stable isotopes. Different elements behave differently when rock partially melts; some tend to concentrate in the melt, while others remain behind.

Delving into the samarium and neodymium of volcanic and altered sedimentary Nuvvuagittuq rock, O'Neil and his colleagues found isotopic signs that the rock could represent the oldest section of crust on Earth. Geochemists had already found rock in Greenland that, according to its isotopes, had been derived from the earliest mantle rock. But by 4.3 billion years ago, that mantle rock had partially melted to yield crustal rock, so researchers had fin-

gered "protomantle" by analyzing Greenland rock derived from it. But where was the "protocrust" that must have been formed as the protomantle formed?

O'Neil and colleagues think they now have such protocrust in Quebec. The Nuvvuagittuq rock has the opposite neodymium isotope signature of the Greenland rock's protomantle. Either this rock is a 2-kilometer-long sliver of protocrust resembling today's iron-rich ocean crust, or it was derived from such protocrust. "That's a first," says geochemist Albrecht Hofmann of the

Older than dirt. Rocks by Hudson Bay may date back to when Earth first separated its primordial stuff into mantle and crust.

Max Planck Institute for Chemistry in Mainz, Germany. "It's an heroic effort" to measure the subtle isotopic variations involved.

The group goes further, drawing on the clocklike radioactive decay of samarium-146 to calculate an age of formation of the Nuvvuagittuq rock of about 4.3 billion years. If accurate, that age would mean they have the protocrust itself, not just something derived from it. That rock would be the oldest rock known, approaching the age of individual zircon mineral grains from western Australia that tell of a wet and weathered world soon after Earth's origin. The new age "is exciting," says geochemist Mukul Sharma of Dartmouth College, but uncertainties remain about details of the rocks' formation that bear on its isotopic age. "There's a lot more work that needs to be done," he says, before a new world's most ancient rock can be crowned.

–RICHARD A. KERR

EPA Nixes Perchlorate Standard

After a multiyear bureaucratic fight, the U.S. Environmental Protection Agency (EPA) has decided not to regulate a toxic rocket fuel component leaching into the nation's drinking water. The proposed ruling explains that requiring the cleanup of perchlorate, which is polluting areas near U.S. military sites, would provide no "meaningful opportunity for health risk reduction." It's a controversial decision. "There is substantial evidence that this chemical needs to be regulated," says toxicologist Melanie Marty, chair of EPA's child health advisory committee. In a 2006 letter to EPA, Marty cited studies that suggest current perchlorate levels at hundreds of U.S. sites could "result in exposures that pose neuro-developmental risks" to infants.

–ELI KINTISCH

Scientists Go Nano a Mano

Last week, the U.S. Environmental Protection Agency and the National Science Foundation jointly funded two centers to track the environmental implications of nanomaterials for 5 years. A $24 million center led by the University of California, Los Angeles, will perform cell-based studies to find materials that pose the greatest potential risks. The other center, awarded $14 million and led by researchers at Duke University in Durham, North Carolina, will track the effects of nanomaterials on organisms as they move through tightly controlled ecosystems in labs. Andrew Maynard, a nanotechnology expert with the Woodrow Wilson International Center for Scholars in Washington, D.C., says the work "is an important step." But he would prefer to see a "robust federal risk research strategy" to systematically evaluate dangers from all potential nanomaterials. –ROBERT F. SERVICE

Gray Wolf Regains Protection

The U.S. Fish and Wildlife Service has put the gray wolf back on the endangered species list. The listing conforms to a U.S. district court order issued in July (*Science*, 25 July, p. 475). There had been speculation that the government might appeal the ruling, which came after the Natural Resources Defense Council and other groups challenged a February decision by the agency to delist the Northern Rockies wolves. Now scientists with the council say they're cautiously optimistic about the wolf's chances. Two thousand wolves roam the region, they calculate; more than 2500 are needed for proper genetic mixing.

–ELI KINTISCH

JOHN BEDDINGTON INTERVIEW

U.K. Science Adviser Makes His U.S. Debut

WASHINGTON, D.C.—Last week, during his first visit to the United States as the U.K. government's chief scientific adviser, John Beddington sat down with *Science*'s news editors to discuss topics as varied as food, fuel, and physics. Nine months into his job, Beddington has adopted a lower profile than his headline-grabbing predecessor, David King. A population biologist at Imperial College London, Beddington has specialized in applying biological and economic tools to questions of natural resource management, particularly fisheries (*Science*, 22 June 2007, p. 1713). He's no stranger to politics, having advised the British government, the European Commission, the United Nations Environment Programme, and its Food and Agriculture Organization. Now Beddington must answer questions from the prime minister and Cabinet, as well as coordinating the science advice in all government departments and chairing a number of committees. The following excerpts from his interview were edited for brevity and clarity.

—DANIEL CLERY

Online
sciencemag.org
Read an extended interview on *Science*NOW.

Q: If you could put one file in the new U.S. president's in-tray, what would it be?

J.B.: The message I would probably want to give is the intimate connection between the issues of climate change, food security, energy security, and water security. These issues need mixed approaches; they need a mix of both science and engineering. These issues are tremendously important because they are going to come quite quickly. The sort of demand increases that are to be expected from urbanization, movement out of poverty, and population growth are quite dramatic, on a time scale of only a couple of decades.

Q: One of David King's goals was to increase the use of science advice across all government departments. Is that job done?

J.B.: I've done a number of things that are slightly different from David. Every 6 weeks, all of the Chief Scientific Advisors of the major departments dealing with science meet with me and with each other. We form subgroups: One is dealing with climate change and food security issues and another is going to be dealing with infectious diseases. That's a good bit of networking. In addition, this group is now meeting with the chief executives of the research councils every 3 months. You now have a network of essentially everybody who's funding government science meeting on at least a 3-monthly interval. A real community is now starting.

Q: David King took a very public stance, putting advice into the public domain even when he disagreed with the government. What approach do you favor?

J.B.: The key thing is that if there's an issue, it needs to be raised. The one that I raised very early on in my tenure was the issue of food security, which I felt had been quite seriously neglected, and the related issue of biofuels. In my first speech [as chief scientific adviser], I raised these issues. Very substantial increases in food prices shortly followed and [there was] a very quick reaction by the prime minister, who raised the issue of food security at the G8 Summit the following summer. Some issues are better raised involving the media and the public at large; others are better talking behind the scenes.

Q: On biofuels, your concern was the competition for arable lands?

J.B.: When I first raised [the issue], I made the point that some biofuels were being produced by cutting down rainforests or using permanent grassland, which has a negative effect on greenhouse gas emissions. So you
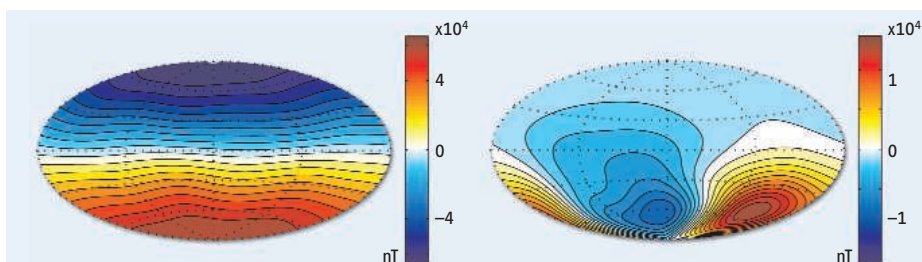
Skewed. Uneven heating of a core producing a normal magnetic field (*left*) concentrates the field (*right*).

GEOPHYSICS

Solid Rock Imposes Its Will on a Core's Magnetic Dynamo

Mariners have been navigating by Earth's magnetic field for centuries. Seismologists detected the fluid-iron core that generates the magnetic field a century ago. But geodynamicists still struggle to understand exactly how the churning of the core's fluid iron generates the field inside Earth. One secret, according to two papers in this issue of *Science*, may lie in the far slower roiling of the solid rock overlying a planet's core. The authors draw on magnetic fields long frozen into the rocks of Earth and Mars to understand how motions in the solid rock can shape a planet's magnetic field.

Here on Earth, the frozen fields link the deep-seated magnetic field to plate tectonics at Earth's surface. On page 1800, paleomagnetist Kenneth Hoffman of California Polytechnic State University in San Luis Obispo and geochronologist Brad Singer of the University of Wisconsin, Madison, draw on magnetic fields locked into lavas as they solidified in Germany and on Tahiti since 780,000 years ago. Five times during a 200,000-year interval, Earth's magnetic field weakened for thousands of years as if it were about to switch its north and south poles, only to return to full strength without reversing. During each such excursion, magnetic field lines that had been pointing in the usual direction—roughly toward the geographic poles—swung around as if one pole were someplace in Eurasia and the other around western Australia.

That pole pattern during ancient excursions has a familiar look, Hoffman and Singer note. Mathematically remove today's powerful, axially aligned dipole field—the sort produced by a bar magnet—from Earth's normal field, and the remaining complex but weak field would skew the pole positions in

CREDIT: S. STANLEY ET AL., SCIENCE

don't want to be doing that. I think that the [U.K. government's] Gallagher Report indicates that there's some need for caution on the development of biofuels within the U.K. and Europe. It's a complicated issue. The information that is available to make a comprehensive assessment of the implications of biofuels is quite inadequate.

Q: You have said that the world needs to dramatically increase food production, using less water than is used today. Will the world need to embrace GM technology?
J.B.: Population growth and the increase in wealth implies something like a 50% increase in food demand by 2030. At the same time, the proportion of the population that lives in an urban environment will go up from about 47% to 60%. That means there's going to be some real problems for agriculture. Essentially, about 70% of available freshwater is used by agriculture. There's going to be competition [for water] between urban communities and agriculture relatively close to urban communities. I'm worried about that.

GM is not going to be the only answer. The knowledge of the plant genome is going to be absolutely critical to improving agricultural production. GM is only one of the

British advisory. John Beddington warns that increasing energy, food, and water demands are vital security issues.

techniques that can be used; marker-assisted breeding could be used equally well.

Q: The U.K. government is falling behind its own targets for reducing greenhouse gas emissions. How should it catch up?
J.B.: There's some interesting work that's being done by the government's new Climate Change Committee, which is going to be reporting in December, that is going to answer those questions very specifically. The Energy Technology Institute, funded jointly by industry and government, is looking at operational

scale inputs to a whole series of green engineering technologies to address these problems. The big [initiative], which everybody really needs to be addressing, is CCS [carbon capture and storage]. And that really needs very serious investment.

Q: At U.K. universities, many physics and chemistry departments have closed because of declining student numbers [*Science*, 12 September, p. 1428]. Should the government intervene to support strategic science subjects?
J.B.: I think it's absolutely critical that we make certain the STEM agenda works—science, technology, engineering, and mathematics are the subjects that we desperately need students to take A-levels [high school finals] in and go on to do degrees. There has been a downward trend [in undergraduate STEM enrollment], but I think it is actually starting to reverse. One area that has been very successful in reversing this [overall] downward trend has been the Ambassador Scheme, in which we've got something of the order of 20,000 scientists and engineers going into schools, talking to students about their lives and the problems they're actually facing. Now our commitment is to expand that.

---

just that way. Hoffman and Singer infer that this field, called the nonaxial dipole (NAD) field, was there three-quarters of a million years ago. Ever since then, the pair argues, something must have kept the molten iron of the core swirling in the same pattern to generate the NAD field.

The ultimate stable driving force appears to be plate tectonics. Lots of cold oceanic plates have sunk through the mantle to the top of the core beneath Western Australia. That relatively cold material would cool the underlying core fluids, which would sink, superimposing a weaker but persistent circulation on the one generating the main dipole field. Hoffman and Singer suggest that the field-generating circulations are layered, with the main dipole field generated deep within the outer core and the NAD generated near its top.

Dynamo specialists say this paleomagnetic argument indicates that mantle rock influences the magnetic field, as modern observations had hinted. "It's very likely the

mantle does have a role in the core flow," says geophysicist Peter Olson of Johns Hopkins University in Baltimore, Maryland, "but it's not that easy to say one [field] is shallow and the other is coming from deep."

On Mars, the patches of magnetic field detected from orbit froze into the crust more than 4 billion years ago, not long before the dynamo in the martian core died. Oddly, the patches of field lingering in the northern hemisphere are far weaker than those in the southern hemisphere. The planet's crust also differs between hemispheres. It's thin and low-standing in the north but high and thick in the south. Could the two asymmetries be related? On page 1822, dynamo specialist Sabine Stanley of the University of Toronto, Canada, and colleagues consider the possibility.

In a dynamo computer model, Stanley and her colleagues made the bottom of the mantle colder in the southern hemisphere than in the north. That would be the temperature pattern imposed on the core by a mantle

circulating so as to create the crustal asymmetry: hotter mantle rock slowly rising throughout the northern hemisphere in one great plume—thinning the crust by eroding it—and cooler mantle sinking throughout the southern hemisphere. Researchers have suggested several ways such a mantle circulation might have been created, including a supergiant impact (*Science*, 11 April, p. 165). Once the resulting temperature pattern was imposed on the model mantle, it induced a circulation in the molten core that generated a magnetic field, but almost entirely in the south and only weakly in the north.

Creating a lopsided magnetic field is "a significant accomplishment," says planetary physicist David Stevenson of the California Institute of Technology in Pasadena. But proving that early Mars worked that way will require a better record of early magnetic field behavior, he cautions. Understanding eons-old interactions between the mantle and the magnetic field will take a lot more work.

–RICHARD A. KERR

CREDIT: GREG SCHALER
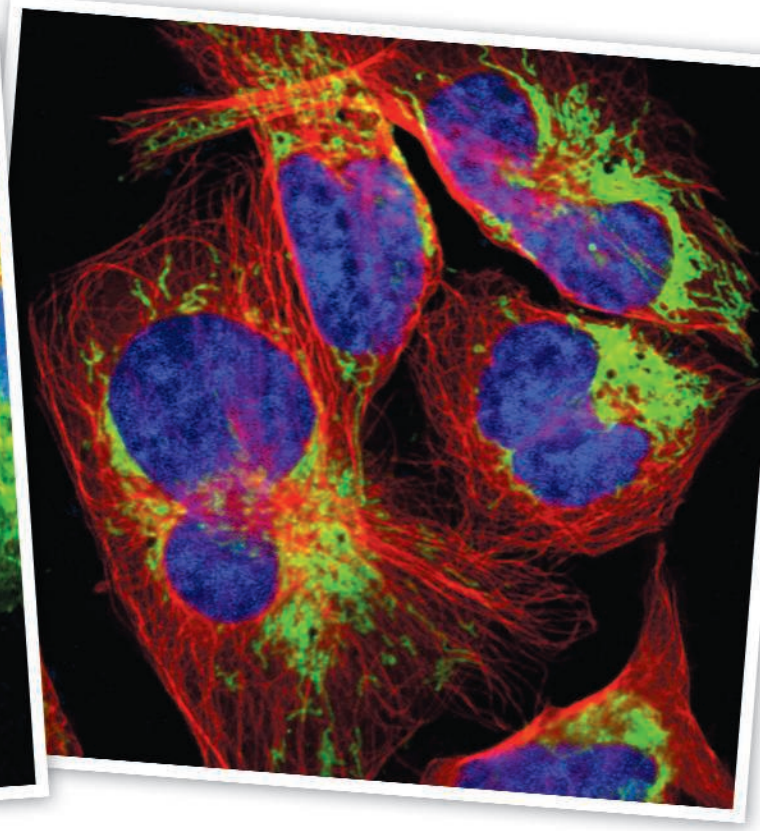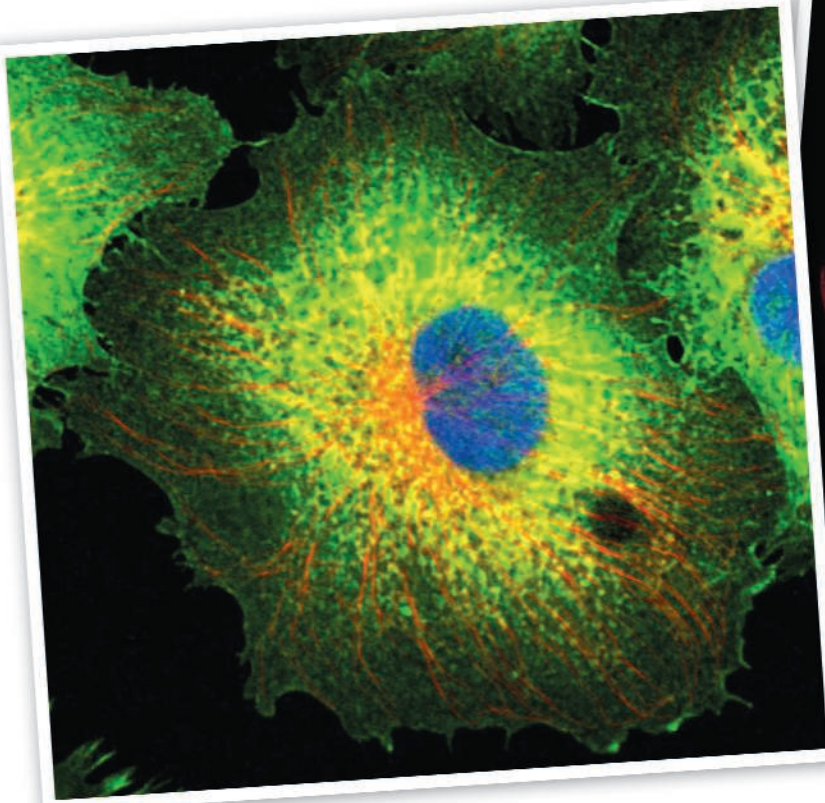
Proteomics Ponders Prime Time

Improved technologies for tracking thousands of proteins at once have spawned talk of a full-scale project to reveal all the proteins in each tissue—but the price tag would be daunting

AMSTERDAM, THE NETHERLANDS—He's too polite to come right out and say it, but Amos Bairoch thinks that much of the data generated by proteomics groups over the past decade is junk. Following the completion of the human genome project, proteomics labs set out to survey all the proteins expressed in different cells and tissues, in essence, putting meat on the bone of the genome. Mass spectrometers and other tools turned out gigabytes of data that purported to identify large numbers of proteins and fed them to Bairoch, who heads Swiss-Prot, a massive database that houses the latest findings on proteins of all stripes. Today, most of those data are ignored, Bairoch says, because the readings were too imprecise to make positive identifications. Throughout the years, many casual observers of the field dismissed proteomics as a waste of time and money. "People thought [the technology] was ready 10 years ago. But they didn't see good results and got disenchanted," Bairoch says.

Today, however, Bairoch's databases and others like them are filling up with terabytes of information that he calls "much better." The upshot: Proteomics is finally coming of age. With the help of better instrumentation and refined techniques, the top proteomics labs can identify and quantify more than 6000 distinct proteins from individual cells and tissues at a time. Now that these labs can cast such a wide net, many proteomics researchers say the time is ripe to undertake a full-scale human proteome project (HPP) to survey the landscape of proteins present in dozens of different human tissues. If successful, such a project would reveal which proteins are actually expressed in different types of cells and tissues, and at what levels, and the network of proteins they communicate with. That knowledge could offer researchers innumerable insights into how organisms convert their genetic blueprint into life and perhaps lead to breakthroughs in biology and medicine. "We are at the point where we can talk about doing this in 8 to 10 years," says Mathias Uhlen, a microbiologist and proteomics expert at the Royal Institute of Technology in Stockholm, Sweden.

It's not just talk. Uhlen and other proteomics leaders gathered here last month to weigh plans for an HPP and to sound out representatives of science funding agencies that would need to pony up the hundreds of millions—if not billions—of dollars needed to pull it off. Most of the responses suggested that tight science budgets make a new mega-sized international science project unlikely anytime soon. Nevertheless, even without a coordinated international HPP, the field is moving so fast that "it's happening already," says Matthias Mann, a proteomics expert at the Max Planck Institute of Biochemistry in Martinsried, Germany.

Spotted history

Many researchers probably assume an international proteome effort started years ago. The availability of the human genome sequence in 2001 told researchers how many proteins are likely to be out there and the exact sequence of amino acids they should look for. The race was on, amid plenty of hype. "Everyone was interested in proteomes," says Mann.

But there were problems, lots of them. For starters, proteins are chemically far more heterogeneous and complex than DNA and RNA. It was relatively easy for researchers to

Revealing. Fluorescent antibodies flag the locations of different proteins in cells, offering clues to those whose functions are unknown.

create a single, robust, and standardized sequencing technology to decode the genetic blueprint of humanity. But no single machine could tell researchers everything they wanted to know about proteins. Worse, although each cell contains the same complement of genes, the abundance of different proteins varies widely. One milliliter of blood, for example, contains about 1 picogram of cell-signaling molecules called interleukins and about 10 billion times that amount of a protein called serum albumin. Such plentiful proteins can mask the signals of their rare brethren.

Still, the lure of proteins was undeniable. Whereas genes are life's blueprint, proteins are the bricks and mortar from which it is built. Identify a critical protein in a disease process, and it could serve as a target for a multibillion-dollar drug to fight diabetes or heart disease. Fluctuations in the amounts of some proteins could serve as "biomarkers" to alert doctors to the onset of cancer or Alzheimer's disease. In the early part of this decade, companies flocked to the field, raising and spending hundreds of millions of dollars. But it quickly became clear that the technology was immature. After several years of trudging down blind alleys, most of the companies that were formed to hunt for biomarkers and drug targets were either folded or merged out of existence (see sidebar, p. 1760).

The news wasn't much better in academia. Take an early example from the Human Proteome Organisation (HUPO), which was launched in 2001 to coordinate international proteomics research and bring order to the unruly field. In 2004, HUPO launched its Plasma Proteome Project (PPP) to survey blood proteins and propel the search for candidate biomarkers. HUPO sent identical blood samples to research groups around the globe, each of which conducted its own analysis with its own homegrown version of the technology. "It was a big disaster," says John Yates, a chemist and mass spectrometry (MS) expert at the Scripps Research Institute in San Diego, California. "There was no quality control. Then the data came back, and it was just a mess," he says.

Unfortunately, PPP and other early efforts raised expectations that they would produce a shortcut for finding novel biomarkers for a wide variety of diseases. "The plasma proteome [project] made the search for biomarkers look like a slam dunk," says Jan Schnitzer, who directs the vascular biology and angiogenesis program at the Sidney Kimmel Cancer Center in San Diego. "But it hasn't delivered." That failure and the failure of proteomics as a whole to deliver on its promise, Uhlen adds, "is a history which is still haunting us."

HUPO has since promoted uniform standards for everything from how to collect and process blood and tissue samples to the proper methodologies for screening them and analyzing the data. And PPP is now taking a more targeted approach to discovering proteins.

The standards have helped, but they haven't solved all the problems. A study last year compared the ability of 87 different labs to use MS to identify correctly 12 different proteins spiked into an *Escherichia coli* sample. No lab got them all, and only one correctly identified 10 of the 12, says Thomas Nilsson, a proteomics researcher who splits his time between Göteborg University in Sweden and McGill University in Montreal, Canada. In a follow-on study completed this year, only six of 24 labs correctly identified 20 spiked proteins. "That again is quite depressing," Nilsson says. "So what are the chances [for success] of high-throughput proteomics as a distributed effort?" Nilsson asked attendees in Amsterdam.

Perhaps surprisingly, Nilsson says he thinks they are decent. This year's study, he explains, shows that most errors in MS-based analyses arise not because the technology can't spot the proteins researchers are looking for but because software programs often misidentify them.

A big part of the problem, says John Bergeron, a proteomics expert at McGill University, rests with simple statistics. To identify proteins using MS, researchers first chop a sample of proteins into smaller fragments called peptides. Those peptides are fed into a mass spectrometer, which ionizes them and shoots them through a chamber. The time it takes for the ions to "fly" through the chamber reveals the atomic weight of the peptides, which in turn reveals their identities. Computer programs then compare them with a full list of the organism's genes, which code for those peptides and their proteins. If a peptide matches the protein code in only one gene, it is a hit and it is a unique identifier of the protein.

The problem is that not all peptides are successfully ionized in each experiment, so some don't enter the chamber. Even if the same lab runs a sample of proteins through the machine twice, Bergeron says, 33% of the proteins identified will appear to be different between the two runs. To minimize such sampling error, MS labs now typically

> *"The biology community at large has to show they really need this. If they can't, why should they fund this?"*
>
> —AMOS BAIROCH, SWISS-PROT

Pacesetter. Thanks to better mass spectrometers and software, researchers such as Matthias Mann (*inset*) can now identify thousands of proteins in a single experiment.

CREDITS: MAX PLANCK INSTITUTE FOR BIOCHEMISTRY

run samples through their machines as many as 10 times. Today, MS groups also look for more than one unique peptide to confirm the identity of a protein. Those changes, together with other emerging standards, show that "these are problems that can be addressed," Schnitzer says.

A new approach

Such successes are also convincing proteomics leaders that the technology is mature enough to go after a full-scale HPP. Although details remain in flux, the generally agreed-upon plan is to identify one protein for each of the estimated 20,400 human genes. Bairoch reported at the meeting last month that Swiss-Prot has logged what is currently known about each gene, such as the primary proteins a particular gene produces and their function. Proteins for about

half of the genes have never been seen, Bairoch stated.

There are far more proteins than genes, because proteins can be spliced together from multiple genes, and once synthesized, they can later be cut down in size or modified with other chemical groups. Trying to find all those variants in all tissues is a task that will likely take decades, Uhlen says. Sticking to one protein for each gene provides a defined endpoint to the project and would create a "backbone" of all human proteins that can be continually fleshed out.

Another possible goal is to create one antibody for every protein in HPP. Because antibodies typically bind to one target and nothing else, researchers can use them to fish out proteins of interest and track their locations in cells and tissues. That would offer clues to the functions of the thousands

of proteins for which little is known. Uhlen and colleagues in Sweden launched just such a global antibody project in 2005. And in Amsterdam, Uhlen reported that the catalog now contains more than 6000 antibodies against distinct human proteins, more than one-quarter of the complete set. At the current rate of new antibody production, Uhlen says his team will finish the task in 2014. More money, he says, would undoubtedly speed the effort.

A third project would track which proteins "talk" to one another. To find a protein's partners, researchers create thousands of identical cell lines and insert into each one a chemical tag linked to a different protein. They can use the tag to pull that protein out of the cell at a specific point in its life cycle, along with any other proteins, bits of RNA or DNA, or a metabolite that it is

Will Biomarkers Take Off at Last?

One much-heralded application of proteomics—detecting proteins that are markers for specific diseases—has long been a dream deferred. "It has been extremely difficult to find those proteins that are biomarkers," says Ruedi Aebersold, a proteomics expert at the Swiss Federal Institute of Technology in Zurich, Switzerland, and the Institute for Systems Biology in Seattle, Washington. But after years of disappointments, proteomics researchers say they're cautiously optimistic.

When proteomics caught fire earlier this decade, scientists hoped that mass spectrometry (MS) and other technologies would help them sift through the thousands of proteins in blood and other body fluids to identify a rare protein that indicated the presence of a disease. Researchers could then use these biomarkers to spot diseases in their formative, treatable stages. But the demise of several companies, such as GeneProt and Large Scale Biology, that jumped into the field revealed that nailing down biomarkers is harder than it sounds.

One problem is that blood—the most common hunting ground—is difficult to work with. Levels of different proteins in blood vary by 10 orders of magnitude, and the abundant proteins often mask the presence of rare ones. Unfortunately, mass spectrometry, the best tool for casting a wide net to search for proteins, hasn't been sensitive enough to spot the rare ones. "Most clinically used biomarkers are at nanogram [per milliliter] levels or below," Aebersold says. At the meeting, Aebersold reported a new strategy for targeting protein fragments called N-linked glycopeptides, which commonly make up cell-surface receptors and thus are more likely to be shed into the blood. This targeting allowed Aebersold's team to spot proteins down to nanogram-per-milliliter levels and thereby track them to look for possible links to diseases. Aebersold says he's hopeful that similar, more focused, studies will improve prospects for the biomarker hunters.

Better instrumentation won't solve all the problems. Techniques such as MS that survey thousands of different compounds inevitably
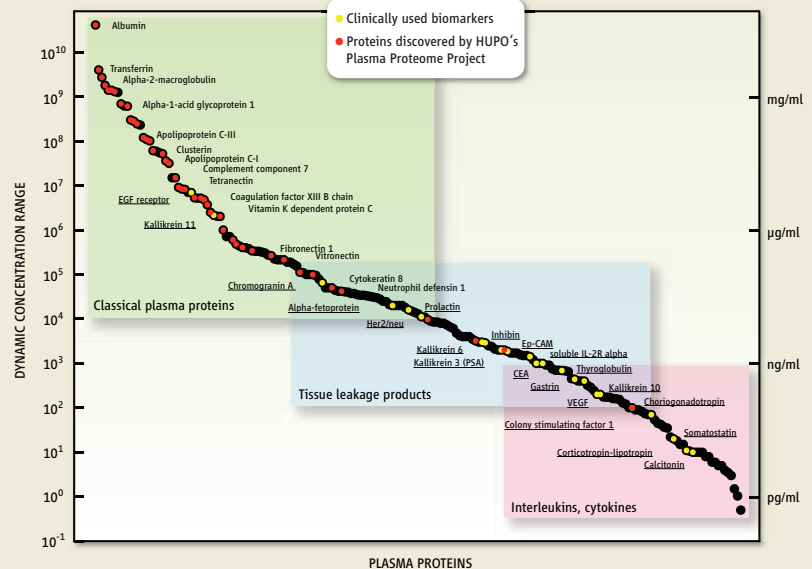
Needles in a haystack. The abundance of different proteins in blood varies by more than 10 orders of magnitude. Most commercially used biomarkers (yellow dots) are present in only minute quantities in blood, below the level at which most proteins are detected (red dots).

turn up false positives: proteins that change their abundance in lockstep with a disease just by chance. That means candidate biomarkers must be validated through clinical trials, which can cost tens of millions of dollars—and most of them fail. "To be accepted by [regulatory] agencies, it's almost as costly as developing a new drug," says Denis Hochstrasser, the director of laboratory medicine at Geneva University Hospital in Switzerland. Because diagnostics companies, unlike drugmakers, typically can't charge lofty premiums for their new tests, they have less incentive to develop biomarker tests. Michael Snyder, a proteomics expert and cell biologist at Yale University, says that despite these challenges, he's hopeful that improving proteomics technologies will generate novel biomarkers—"just not on the same time frame as people thought."

–R.F.S.

bound to. Bioinformatics experts can then weave together the partners for each protein to construct a complete communication network of the proteins in the cell.

Such protein-interaction networks have been worked out in exquisite detail in yeast and other organisms. But it has been hard to insert the chemical tags reliably into human cell lines. Over the past decade, however, researchers around the globe have shown that different lentiviruses readily insert tagged proteins into a wide variety of human cells. At the meeting, Jack Greenblatt of the University of Toronto in Canada said he has proposed a project to insert one tagged protein for each of the 20,400 genes, the first step to a complete human proteome interaction map. The project is now under review by Genome Canada, the country's national genome sciences funding agency. Greenblatt adds that working with human cell lines isn't perfect, because these lines are typically made up of non-normal cells that have been immortalized. His group is also performing related studies in mice, which can be grown into adult animals, and the interaction networks can be compared with those found in the human cell lines. Other projects could be added to HPP as funding permits. They could include a catalog of all the modified proteins, such as splice variants and phosphorylated proteins, Bergeron says.

Finding the money

How much will it take to complete the wish list? Opinions vary, but somewhere in the neighborhood of $1 billion is a common guess. Michael Snyder, a yeast biologist at Yale University, thinks that's too little. "This is going to require a bigger budget than that," he says.

Whatever the projection, it was enough to make those with the money blanch. Funding agencies around the world are already collectively spending hundreds of millions of dollars on proteomics technologies and centers. They're also already committed to several international big biology projects such as the International HapMap, the International Cancer Genome Consortium, and the Knockout Mouse Consortium, which are putting the squeeze on tight budgets. "From a funding viewpoint from the U.S. context, now is not the right time," says Sudhir Srivastava, who directs proteomics initiatives at the U.S. National Cancer Institute in Rockville, Maryland. "If this was 5 years ago when the NIH [National Institutes of Health] budget was doubling. …" Srivastava trails off.

Still, Uhlen and others say they are hopeful that funding agencies will keep the field moving quickly. "We don't have to have $1 billion from the start," Uhlen says. "With the Human Genome Project, it took 5 years for the funding agencies to put serious money into it. I don't think we should expect funding agencies to jump on board until we have proven the technology."

To do that and make the cost more palatable, HUPO leaders are mulling a pilot project to catalog all the proteins produced by chromosome 21, the smallest human chromosome, which has 195 genes. Although the cost of such a project isn't known, "I think there almost certainly would be interest," says Roderick McInnes, director of the Institute of Genetics at the Canadian Institutes of Health Research in Ottawa.

At the meeting, proteomics expert Young-Ki Paik of Yonsei University in Seoul, South Korea, said the Korean government is considering funding a similar proposal for a Korean-based pilot project on chromosome 13, the second-smallest human chromosome, with 319 genes. Paik says he and his colleagues have proposed a 10-year, $500 million initiative that is currently being considered by the Korean Parliament. A decision is expected in October. If it is funded, Bergeron says it will be a major boost to the field and could help catapult Korea into the forefront of proteomics.

Some researchers are skeptical of going chromosome by chromosome, however. "In gene sequencing, that approach worked," Bairoch says. "You could separate out the work by chromosome. But it doesn't make sense for proteins. There is no [body] fluid or [tissue] sample organized by chromosome." Ruedi Aebersold, an MS expert with a joint appointment at ETH Zurich and the Institute for Systems Biology in Seattle, Washington, agrees. "I'm not a big fan of going chromosome by chromosome," he says. MS machines, he notes, identify whatever proteins show up regardless of the chromosomes they came from.

Whatever path they take to an HPP, proteomics leaders will need to find true believers beyond those already in the flock. "The biology community at large has to show they really need this," Bairoch says. "If they can't, why should they fund this?" Uhlen, Bergeron, and other HUPO leaders

"We are at the point where we can talk about doing this in 8 to 10 years."

—MATTIAS UHLEN, ROYAL INSTITUTE OF TECHNOLOGY

agree. And they argue that current demonstrations of the technology are starting to build the case.

At the Amsterdam meeting, for example, Mann reported that recent advances in instrumentation and software have enabled his group to identify the complete yeast proteome in one shot—in just a few days. That feat took months of painstaking effort when it was first accomplished by traditional methods 5 years ago. Mann also described the use of a technique his team first reported last year to monitor changes in the yeast proteome, including levels of individual proteins, between two different states. In one example, Mann's team compared yeast cells with a diploid (double) set of chromosomes to cells with the haploid (single) set undergoing sexual reproduction. The study quantified for the first time the suite of proteins that orchestrate sexual reproduction in yeast. Mann says the technique opens the door to studying proteomewide differences between healthy and diseased cells, developing and mature cells, and stem cells and differentiated cells. "There is no end to what you can compare," Mann says. "Every lab can ask these questions."

In an another study, Uhlen reported using his antibodies to track global protein expression in human cells. He and his colleagues have shown that fewer than 1% of all proteins are expressed in only one tissue. That implies, he says, that tissues are differentiated "by precise regulation of protein levels in space and time, not by turning expression on and off." Aebersold also reported that his lab has devised a scheme for detecting proteins expressed at the level of just a single copy per cell.

"These are unbelievable advances, and they show we can take on the full human proteome project immediately," Bergeron says. Not everyone has turned that corner, but Bergeron and others say that they are confident that time is coming soon. As Pierre LeGrain, director of life sciences at the French Commissariat à l'Energie Atomique in Gif-sur-Yvette, sums it up: "Most of us feel the human proteome project is going to happen, though we don't know how."

–**ROBERT F. SERVICE**

ELECTION 2008

Scientists Strive for a Seat at the Table of Each Campaign

When it comes to soliciting scientific advice, Barack Obama welcomes a cast of thousands, whereas John McCain plays it close to the vest

Harold Varmus has met Senator Barack Obama only once. But he's convinced that the Democratic presidential nominee "understands the important role that science must play in tackling the problems we face as a society." To prove it, the president of Memorial Sloan-Kettering Cancer Center in New York City points to the candidate's promise of "sustained and predictable increases in research funding" at the major federal science agencies.

It's no surprise that the politically active Varmus, the 1989 medicine Nobelist and former director of the U.S. National Institutes of Health (NIH), is familiar with Obama's statements on funding basic research: He helped write many of them as chair of a 40-plus-member committee of prominent researchers and educators who are advising the freshman senator from Illinois on science. The panel prepared the candidate's 6000-word response last month to 14 questions posed by a coalition of scientific organizations called Science Debate 2008 (ScienceDebate2008.org). Varmus won't say how much the answers were altered by campaign officials but allows that "we're very pleased with it. His commitment to science is absolutely apparent."

Last week, Obama's Republican opponent, Senator John McCain (AZ), provided equally lengthy answers to the same set of questions. Douglas Holtz-Eakin, who serves as the candidate's point man on many domestic policy issues, including science, health, energy, and the environment, says McCain has contacted experts on issues such as climate, space, and "science in general" but has "no formal structure" for soliciting advice. An economist and former head of the Congressional Budget Office under President George W. Bush, Holtz-Eakin says McCain relies instead on the knowledge acquired during his 26 years in Congress, including 6 years as chair of the Senate Commerce, Science, and Transportation Committee.
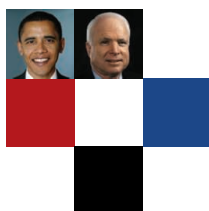
The way the answers were prepared reflects the different management styles of the two campaigns. "Obama has thousands of advisers, and McCain has two guys and a dog," cracks one academic lobbyist who requested anonymity because his organization tries to maintain ties with both camps.

The answers themselves—on research funding, science education, climate change, energy, space exploration, and other issues—reflect different political philosophies. Obama tends to assign government a larger role in tackling those problems—a $150 billion plan for energy independence, for example, and an $18 billion plan to improve education. McCain, in contrast, combines his $30 billion clean-coal program with talk about the need to curb spending and rely on the private sector.

For many U.S. academic researchers, presidential politics comes down to two big issues: getting more money for science and having a seat at the table. The first requires agreement between the president and Congress, however, and any promise to increase research spending could easily be derailed by the Iraq war, an ailing economy, and rising health care and energy costs. That puts a premium on the second issue, namely, the appointment of people who will make the key decisions in the next Administration.

Indeed, three independent panels stuffed with science mavens have recently issued reports[*] emphasizing the importance of choosing an assistant to the president for science and technology soon after the election. They say that person, who would also head the Office of Science and Technology Policy (OSTP), should be part of the president's inner circle and play a major role in vetting appointments to dozens of other key science positions throughout the government.

"We had drafted white papers on several issues, but the presidents [of the three academies] worried that nobody would pay attention to them," says E. William Colglazier, executive officer of the U.S. National Academy of Sciences, which joined with the engineering academy and the Institute of Medicine in issuing a report last week on the appointment process. "They felt it was more important that the next president get very good people into key positions." Working backward, scientists reason that the more interaction with a candidate before election day, the greater the chance that he will act quickly and fill those posts with highly qualified people.

Since declaring his candidacy in February 2007, Obama has welcomed those interactions. He has solicited the views of troves of experts and created a vast network of advisers. "They didn't ask us to take a blood oath," says Varmus, who endorsed Obama with the Democratic nomination still hanging in the

---

Science and the **2008** **Campaign**

Many voices. Harold Varmus (*far left*) and other leaders joined Barack Obama at a June economic roundtable at Carnegie Mellon University.

CREDIT: CARNEGIE MELLON UNIVERSITY

balance. But he says "it's a reasonable assumption" that most of the advisers also support his candidacy.

Varmus's panel, which includes medicine Nobelist Peter Agre and physics Nobelist Leon Lederman, is one of 20 or so advisory bodies. (The Obama campaign declined to provide a number.) Paul Kaminski, a top Pentagon official during the Clinton Administration, is heading up an eight-person group on defense science that is examining work-force, training, and acquisition issues. He'll also be representing Obama next month at a National Academy of Engineering forum on grand challenges, opposite Carly Fiorina, the former CEO of Hewlett-Packard who was once on McCain's list of possible running mates. There's another Obama group on science education, and the membership is overlapping. Kaminski recently joined the science panel, for example, and Lederman also serves on a small group examining science education.

With regard to scientific input in a McCain Administration, Holtz-Eakin promises that McCain will be vigilant in ending what critics have called the Bush Administration's war on science. "He'll restore credibility and transparency" to the process, says Holtz-Eakin, in part by filling all six statutory positions at OSTP. Still, Holtz-Eakin knows that he's addressing a skeptical audience. "You can't convince people that you'll make sure they have access. You have to demonstrate it," he told *Science*.

Convened this summer, Obama's science group has held weekly teleconferences to field questions from the campaign staff and inject into the campaign issues that it feels are important. In preparing answers to Science Debate's 14 questions, the panel's most visible product, members sifted through Obama's past statements, added their own perspectives, and delivered answers to Jason Furman, Obama's director of economic policy, via his deputy, Larry Strickling.

The panel's fingerprints are evident in the nuanced responses that Obama offers. To a question about how basic research would fare in a competition for scarce funds, for example, Obama discusses the declining success rate among applicants for NIH grants, the resulting pressure on young scientists, and the erosion of the agency's buying power after a succession of flat budgets that followed a 6-year doubling from 1998 to 2003. In such an environment, he adds, scientists are less inclined to take risks.

"This situation is unacceptable," he declares, offering as the solution a 10-year, across-the-board budget doubling in the physical and life sciences, mathematics, and engineering.

McCain is less sanguine than Obama about the likelihood of large increases. "I have supported increased funding at DOE [Department of Energy], NSF [National Science Foundation], and NIH for years," he notes in his Science Debate reply, "and will continue to do so." But he warns that "with spending constraints, it will be more important than ever to ensure that we are maximizing our investments in basic research." And his answer omits mention of any numerical goal. In an interview last month on National Public Radio, Holtz-Eakin

Tight team. Douglas Holtz-Eakin (*inset*) has been the chief spokesperson for John McCain on most domestic issues.

said any call for doubling science agency budgets is "a nice, fun number ... that doesn't reflect a balancing of political priorities."

In fact, it's hard to pin down either candidate on how quickly he would like to increase federal funding for basic research. Making a video appearance this month during a cancer research telethon, Obama promised to double the budget of NIH, including the National Cancer Institute, in 5 years. That's twice the rate described in his answers to Science Debate, which came out in August and have become the mantra for campaign surrogates. He also supports the 2007 America COMPETES Act (ACA), which is silent on NIH but which would put NSF and DOE's Office of Science on a 7-year doubling track.

As it happens, those figures are in line with historical trends. Between 1962 and 2003, for example, the NIH budget doubled roughly every 8 years, in current dollars. NSF has seen its budget double every decade for the period from 1970 to 2000.

A statement on McCain's Web site also promises to "fully fund" the provisions of the

COMPETES Act, which authorizes spending levels that have not been met in subsequent appropriations bills. Holtz-Eakin told *Science* that McCain "is on the record as supporting ACA" and that, if elected, his 2010 budget would reflect those targets in the physical sciences. Taking a jab at Obama's expansive promises for increased spending in research and other domestic areas, Representative Vern Ehlers (R–MI) predicts the U.S. research enterprise will be better off under a McCain Administration, despite its more modest promises, because "he's more likely to find the money." But Ehlers, one of three physics Ph.D.s in Congress and a staunch supporter of science, admits that McCain hasn't sought his advice on the topic. (His colleague, Representative Rush Holt (D–NJ), has spoken for Obama, although during a recent interview with *Science* he deferred several questions to the campaign staff.)

Obama's aides and outside advisers play down the discrepancies in Obama's statements on NIH doubling while at the same time perpetuating them. Domestic policy director Neera Tanden, who joined the campaign this summer after many years advising Senator Hillary Clinton (D–NY) on health-care issues, says a 5-year doubling of the NIH budget "is the right thing to do" and that it is needed to keep pace with the rapid advances in the field. Tanden also says the disruptions caused by a stagnant NIH budget after the previous doubling aren't inevitable. "There's no reason to assume you would have another crash landing," she says.

Gilbert Omenn, a professor of medicine and public health at the University of Michigan, Ann Arbor, and a former president of AAAS (which publishes *Science*) who serves on Obama's science advisory panel, acknowledges that the different timetables "are very awkward" and that the candidate's promises "add up to a lot of commitments." But he's confident that Obama "will be able to figure out the best combination of variables to allow for a sustained investment."

In the end, of course, promises are only that. "Remember, it's a campaign, not governance," notes Lederman when asked if his group expects to have an impact on Obama's education policies if he takes office in January. A seat at the table may be a better bet, says Kaminski. "I would expect some of [his defense advisers] to take key positions in his Administration." That is, if they turn out to have bet on the winning candidate.

–JEFFREY MERVIS

AGING

Searching for the Secrets Of the Super Old

More and more people are living past 110. Can they show us all how to age gracefully?

They were born when the years still started with "18." They survived global traumas such as World War I, World War II, and the Great Depression. They didn't succumb to pandemic flu, polio, AIDS, Alzheimer's disease, or clogged arteries. Supercentenarians, or people who've survived to at least age 110, are longevity champions.

Living to 100 is unlikely enough. According to one estimate, about seven in 1000 people reach the century milestone. And at that age, the odds of surviving even one more year are only 50–50, says James Vaupel, director of the Max Planck Institute for Demographic Research in Rostock, Germany. Making it from 100 to 110 "is like tossing heads 10 times in a row."

Researchers are keen to investigate these 19th century holdovers. "If we want to better understand the determinants of longevity, we have to look at the oldest old," says biodemographer Jean-Marie Robine of INSERM's demography institute in Montpellier, France. With Vaupel, he has recently compiled a demographic database of verified supercentenarians from the industrialized countries.

Two other projects, led by researchers on the opposite coasts of the United States, hope to pin down the traits of these survivors by surveying their genomes for longevity-promoting DNA sequences and by autopsying them when they finally die. Ultimately, work on supercentenarians could uncover "a unique [genetic] variation that explains their longevity that can be the subject of drug development," says molecular geneticist Nir Barzilai of Albert Einstein College of Medicine in New York City. Such a discovery might not stretch human life span, but it could make our final years less grueling, suggests Barzilai.

Yet studying supercentenarians is no easy task. Finding these one-in-a-million people is hard enough, and validating their ages can require that researchers become detectives or hire ones.

Come on, how old are you really?

Figures on the number of supercentenarians are shaky. The 2000 U.S. census claimed a total of 1400 living in the country. That number is much too high, says geriatrician Thomas Perls of Boston University School of Medicine, head of the New England Centenarian Study and its new National Institutes of Health–funded spinoff, the New England Supercentenarian Study. Researchers suspect that some of the oldsters included in the tally had already died and that others—or their relatives—were lying about their ages. Drawing on Medicare enrollment figures, two U.S. government actuaries put the number of supercentenarians in the year 2000 at a mere 105. And in 2002, 139 people claiming to be at least 110 were receiving Social Security payments.

"Claiming" is a key word. A crucial part of studying supercentenarians is proving that they were or are their stated age. No biochemical test or medical exam can peg how old somebody is. So researchers often turn to Robert Young of Atlanta, Georgia, a self-taught documents guru who confirms the ages of the world's oldest people for *Guinness World Records*. Young comes across like a veteran insurance adjuster who's seen all the scams. To weed out pretenders,
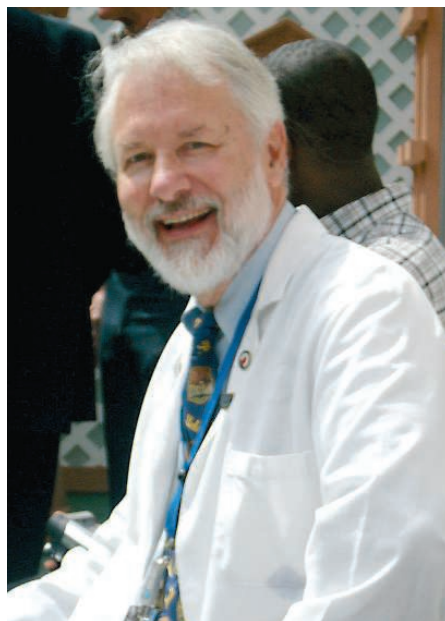
Cutting for clues. L. Stephen Coles leads a group that has performed most of the autopsies on supercentenarians.

he requires three types of verification: proof of birth, preferably a birth certificate; proof of death, if the person is no longer alive; and "continuity" documentation, such as a driver's license or marriage certificate, that shows that the putative supercentenarian is the person listed in the birth record. If candidates or their families can't provide corroboration, Young sleuths through census rolls, school and military records, genealogies, and other types of paperwork.

Using these methods, an organization called the Gerontology Research Group verifies the ages of living supercentenarians and posts a list online (www.grg.org). Young is senior claims examiner for the group, which is headed by L. Stephen Coles of the University of California, Los Angeles, an ob-gyn and computer scientist by training. As of last week, the roster included 10 men and 68 women from 12 countries, ranging up to 115 years old. For reasons that remain murky, most supercentenarians are women. Moreover, of the oldest people ever documented, the majority have been women, including the record-holder Jeanne Louise Calment of France, who died in 1997 at the age of 122.

To obtain a more complete count of supercentenarians for demographic analyses, Vaupel, Robine, and colleagues have dug into national archives, including the records of the U.S. Social Security Administration, to compile lists of candidates in 15 industrialized countries. A team of age checkers then vetted each case. In all, the new International Database on Longevity caches information on nearly 1000 supercentenarians from the past 50 years, although not every country's records span this entire range. The researchers plan to publish a monograph on the database later this year.

But that still won't be the final word. A

Turn of the century. The world's oldest living person is 115-year-old Edna Parker (*right*). Daniel Guzman (*left*), reached 111 before dying earlier this year.

lack of good records in developing nations means that researchers still know little about the numbers of supercentenarians worldwide, says demographer Bertrand Desjardins of the University of Montreal in Canada: "It's anyone's guess how many supercentenarians are living in China."

How to grow old in style

Scientists have gotten a few hints about what keeps centenarians alive for so long—genes associated with a beneficial lipid profile, for example (*Science*, 17 October 2003, p. 373)—but they're just beginning their search for the sources of supercentenarian longevity. Two years ago, Perls and colleagues published the first health survey on these so-called supers, reporting on 32 people between the ages of 110 and 119. "I think it's incredible how well off they are," says Perls. Although almost half of the supers had osteoporosis and almost 90% had cataracts, 41% of them either lived on their own or required only minimal help with tasks such as preparing food, dressing, and bathing. Cardiovascular disease, the leading killer in developed countries, was rare among supercentenarians—only 6% had suffered heart attacks and 13% reported strokes. Diabetes and Parkinson's disease were also uncommon in the group, striking only 3% of the subjects each. Like centenarians, supercentenarians seem to be good at putting off the day when they become disabled, says Perls.

The superseniors deviate from the norm not just in how long they live but in how they die, says Coles, who arranges autopsies of the oldest old as part of his work with the recently established Supercentenarian Research Foundation. Only nine supercentenarians have undergone postmortems—Calment, for example, never agreed to one—and Coles and colleagues have performed six of these procedures, including one earlier this year in Cali, Colombia, on a man who died at age 111.

Coles argues, based on these autopsies, that supers aren't perishing from the typical scourges of old age, such as cancer, heart disease, stroke, and Alzheimer's disease. What kills most of them, he says, is a condition, extremely rare among younger people, called senile cardiac TTR amyloidosis. TTR is a protein that cradles the thyroid hormone thyroxine and whisks it around the body. In TTR amyloidosis, the protein amasses in and clogs blood vessels, forcing the heart to work harder and eventually fail. "The same thing that hap-

pens in the pipes of an old house happens in your blood vessels," says Coles.

Perls and colleagues have also shown that extreme survival runs in supercentenarians' families. Repeating an analysis they did earlier for centenarians, the researchers last year analyzed life spans of the siblings and parents of supercentenarians from the United States. The team compared the relatives' longevity with that of people born in the same year. Brothers of supers gained about 12 to 14 years over their contemporaries, whereas sisters outlasted their counterparts by about 8 to 10 years. A family connection doesn't mean that only genes are responsible for supercentenarians' great age, Perls cautions. Everything from diet to exercise habits can also run in families—the analysis can't distinguish between genetic and environmental factors.

But Perls's current work might. He and his colleagues have collected blood samples from 130 authenticated supercentenarians and have sequenced DNA from 100 of them. As early as this fall, the team could be ready to submit a paper on gene variants that might be stretching supercentenarians' lives, he says. Moreover, because the research team also has data on the supers' past health and lifestyles, it might be able to statistically tease apart environmental and genetic influences on the oldsters' life spans.

The Supercentenarian Research Foundation has similar ambitions. In addition to the autopsies this nonprofit group of doctors and researchers has conducted, Coles and his colleagues have obtained a few blood samples and plan to start collecting more next year. However, the effort, which is operating on donations, won't have enough money to sequence DNA from the samples. They will go into the freezer, but Coles says the shrinking costs of sequencing technology should soon make reading the DNA affordable.

The two projects will be sharing Young's age-checking services but nothing else. Perls says he declined to collaborate with Coles's group in part because some of its members are involved in so-called antiaging medicine, whose practitioners claim to be able to alleviate time's ravages with treatments such as injections of human growth hormone (HGH) (*Science*, 8 February 2002, p. 1032). The

rationale is that the hormone's blood levels normally dwindle as we age. But Perls has blasted this off-label use of the hormone—it's only approved for children with stunted growth and adults with pituitary tumors or other rare conditions—as not only unproven and potentially unsafe but also illegal in the United States.

Working quietly outside that fray, Vaupel and colleagues plan to use their new database to answer a question that's been nagging

Saved by a SNP? Tom Perls (*right*) and colleagues are scanning DNA from people like 110-year-old Mary Marques for longevity clues.

demographers and actuaries: Do the odds of dying in a given year, which rise relentlessly for most of adult life, taper off in the most senior seniors? Demographers want to determine whether the death rate stabilizes so they can test their models of mortality, whereas actuaries need the answer to help governments refine budgets for health care and pensions. If mortality does peak or even begin to decline in very old age, it could mean that people who live past 110 really are super, stronger than the rest of us, Vaupel says.

If centenarians are any guide, researchers will find that supercentenarians have varying backgrounds, lifestyles, and genetic profiles. But as Robine notes, they share one factor: luck. Calment provides a prime example. She outlived her husband, daughter, and grandson. They died from non-aging-related causes—the husband from food poisoning, the daughter from pneumonia, and the grandson in a car accident. So if you hope to reach the big 110, keep a rabbit's foot handy.

–MITCH LESLIE

2008 Visualization Challenge

THE ILLUSTRATION ON THE COVER OF THIS WEEK'S ISSUE OF *SCIENCE* IS A MERGER of science and art, with a good helping of whimsy. Turn to page 1772, and you will see that it is part of a larger illustration depicting elements of the natural world, put together to resemble the Mad Hatter's tea party in Lewis Carroll's *Alice in Wonderland*. It's a magnet for attracting young children into the world of science, a quality that helped it win the first-place award for informational graphics in this year's International Science & Engineering Visualization Challenge.

Online
sciencemag.org

Slideshow and podcast interview with finalist judge Alisa Machalek.

For the past 6 years, *Science* and the U.S. National Science Foundation (NSF) have cosponsored annual challenges to encourage cutting-edge efforts to visualize scientific data. We have been supporting these competitions because we firmly believe that bringing data to life visually will be increasingly important in a world in which images, from traditional media to YouTube, are a primary means of communication. New ways of conveying scientific data will be essential not only for increasing public understanding of science and engineering but also for improving communication across scientific disciplines.

This year, we received 181 entries from 20 U.S. states and the District of Columbia and 20 countries. A committee of staff members from *Science* and NSF screened the entries, and an outside panel of experts in scientific visualization reviewed the finalists and selected the winners. The winning entries appear on the following pages.

We encourage you to submit applications for next year's challenge, details of which will be available at www.nsf.gov/news/special_reports/scivis/index.jsp, and to join us in celebrating this year's winners.

Susan Mason of NSF organized this year's challenge. Rachel Zelkowitz of *Science*'s news staff wrote the text that accompanies the images in this special section, and Martyn Green and Tara Marathe put together a special Web presentation at www.sciencemag.org/vis2008.

JEFF NESBIT, DIRECTOR, OFFICE OF LEGISLATIVE AND PUBLIC AFFAIRS, NSF
MONICA BRADFORD, EXECUTIVE EDITOR, *SCIENCE*

JUDGES

MICHAEL KEEGAN
Michael Keegan Design
Reston, Virginia

GARY LEES
Johns Hopkins University School of Medicine
Baltimore, Maryland

ALISA ZAPP MACHALEK
National Institute of General Medical Sciences
Bethesda, Maryland

MALVINA MARTIN
National Geographic Television
Washington, D.C.

ROBERT PATTERSON
National Center for Supercomputing Applications
University of Illinois, Urbana-Champaign

Photography

FIRST PLACE
THE GLASS FOREST

Mario De Stefano, The Second University of Naples

DIATOMS ARE TINY CREATURES, BUT THEY PLAY A BIG ROLE IN CREATING BREATHABLE air; they produce as much as 40% of the world's oxygen. They can also possess an ethereal beauty, as the winning entry, "The Glass Forest" by Mario De Stefano of The Second University of Naples, Italy, shows. De Stefano used a scanning electron microscope to capture these images of the diatom *Licmophora ehrenbergii* from the Mediterranean Sea off the coast of Italy.

The title refers to the fact that diatoms are unique in using silica to build their cell walls and to the interactions between the diatom and its host, which echo the interplay of a forest. Each green, triangle-shaped diatom measures about 30 micrometers across. There are about 100,000 species of diatoms, the largest of which can reach up to 2 millimeters in length.

The micrograph shows *L. ehrenbergii* clinging to the marine invertebrate *Eudendrium racemosum*, in brown. De Stefano intended the image to depict the complex interactions that occur among organisms on the microscopic scale. "This is the study of a community, … the same community as you would find in a rainforest."

Panel of finalist judges member Malvina Martin praises "The Glass Forest" for its arresting beauty: "This is an invisible piece of our world that plays such an important role. Everyone was pretty awed by that."

HONORABLE MENTION
STRING VIBRATIONS

Andrew Davidhazy,
Rochester Institute of Technology

A TWIST OF THE FINGERS SENDS A STRING INTO a feverish dance. Photographer Andrew Davidhazy of the Rochester Institute of Technology in New York set out to photograph that dynamic by attaching a tiny motor to a cotton string. But he got overzealous in powering up the motor, forcing an atypical torque in the string. "I happened to overspin the string and all of a sudden, the picture was more exciting," he says. Davidhazy used a Canon digital camera to document the movement. The total exposure time for "String Vibrations" was about 2 seconds, during which the string spun some 10 to 20 times, the photographer says.

SQUID SUCKERS: THE LITTLE MONSTERS THAT FEED THE BEAST

Jessica D. Schiffman and Caroline L. Schauer, Drexel University
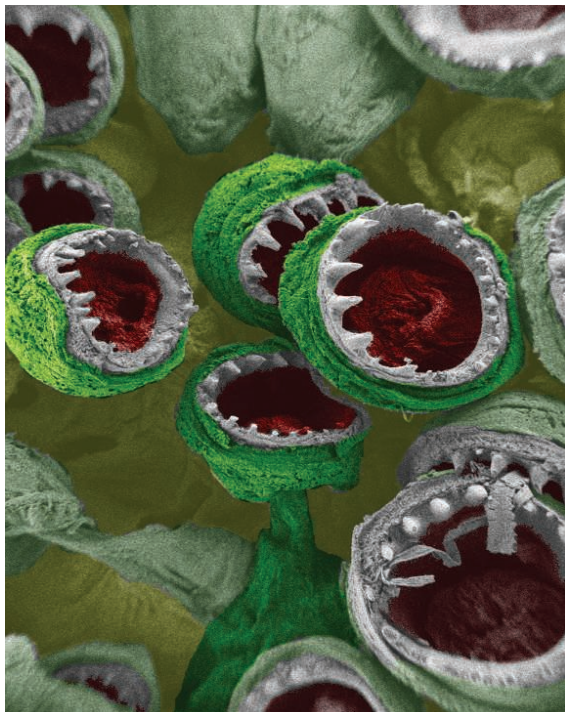
CRUNCH. THE SATISFYING SOUND OF A CRUSHED COCKROACH COMES from the destruction of its chitin-based exoskeleton. The white, fang-like circles in this electron micrograph of squid suckers are also chitin, but they are not so easily crushed. Their scant 400-micrometer diameter belies the true power of the suckers. A squid uses them to latch onto prey and force the unfortunate creature to its beak, where it is readily slurped down. "They're just tiny things, but they really keep the beast alive," says Jessica Schiffman, a doctoral student in material science engineering at Drexel University in Philadelphia, Pennsylvania. She compiled the image while researching chitin properties in the lab of Caroline Schauer. The iconic film *Little Shop of Horrors* inspired the color scheme, she says.
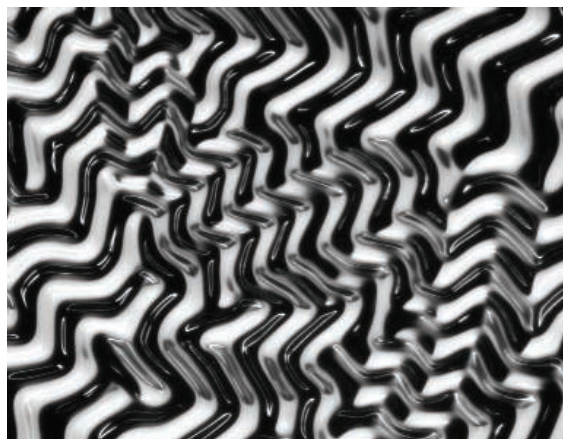
POLYMAZING

Ye Jin Eun and Douglas B. Weibel,
University of Wisconsin, Madison

THE COMBINATION OF POLYETHYLENE GLYCOL (PEG) AND polydimethylsiloxane creates what Ye Jin "Jenna" Eun calls a "sandwich of polymers." But when the University of Wisconsin, Madison, doctoral student added water to her creation, it was clear the union of these polymers didn't emulate peanut butter and jelly's happy marriage. PEG wants to expand when it encounters water, but the stiffer polyethylene copolymer won't permit it. So PEG stretches vertically instead, creating the hills and valleys seen here. The contortions make the polymer combo unusable for the original purpose: mounting cell samples. But Eun, who works under the direction of biochemist Douglas Weibel, says the result of the failed experiment was so beautiful, she photographed the image. She used a Zeiss stereoscope and a Nikon CCD camera.

2008 VISUALIZATION CHALLENGE
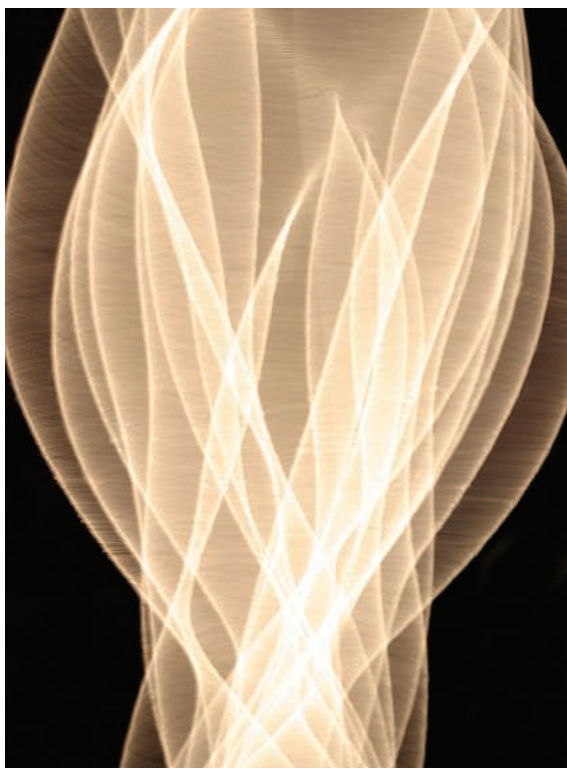
Illustration

ZOOM INTO THE HUMAN BLOODSTREAM

Linda Nye and the Exploratorium Visualization Laboratory,
The Exploratorium

A DISCUSSION OF THE HUMAN CIRCULATORY SYSTEM TYPICALLY BEGINS AND ENDS with the heart. But in this illustration, the team manipulates perspective to show the relationship between the tiniest oxygen atom and the comparatively giant organ. Jennifer Frazier, who directed the project by San Francisco's Exploratorium, says her team used a common technique in landscape paintings to fit multiple scales into a single image. "Something very large can appear small because it's on the horizon, and something very small can appear large because it's in the foreground."

In the image, illustrated by artist Linda Nye, a human heart is in the background and a viewer's eye follows the artery downward to an interior view of the bloodstream in the foreground. The magnification at each level of the image increases 10-fold to show red blood cells and even the oxygen atom within a heme group with colorful clarity.

The goal is to show the interactions between the macro and micro, Frazier says: "The system requires multiple scales to make things happen." The image is meant for display within museums, and its appropriateness for that audience impressed judges, says panel of finalist judges member Alisa Machalek: "It really accomplished the goals of using art to explain science." Fellow judge Michael Keegan adds, "It's just a good way of presenting that macro-micro situation where tiny parts of the circulatory system contribute to the life of the whole body."
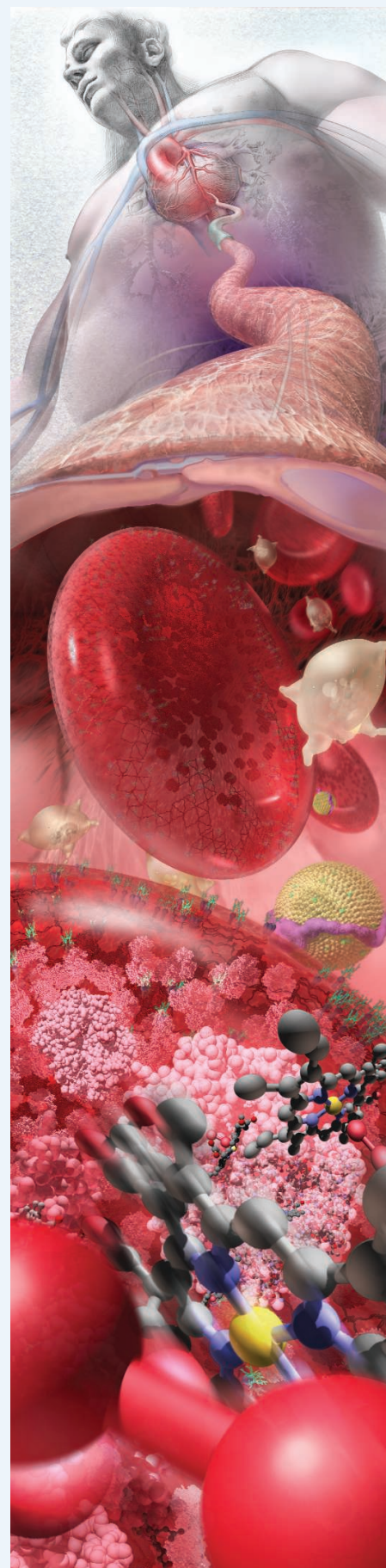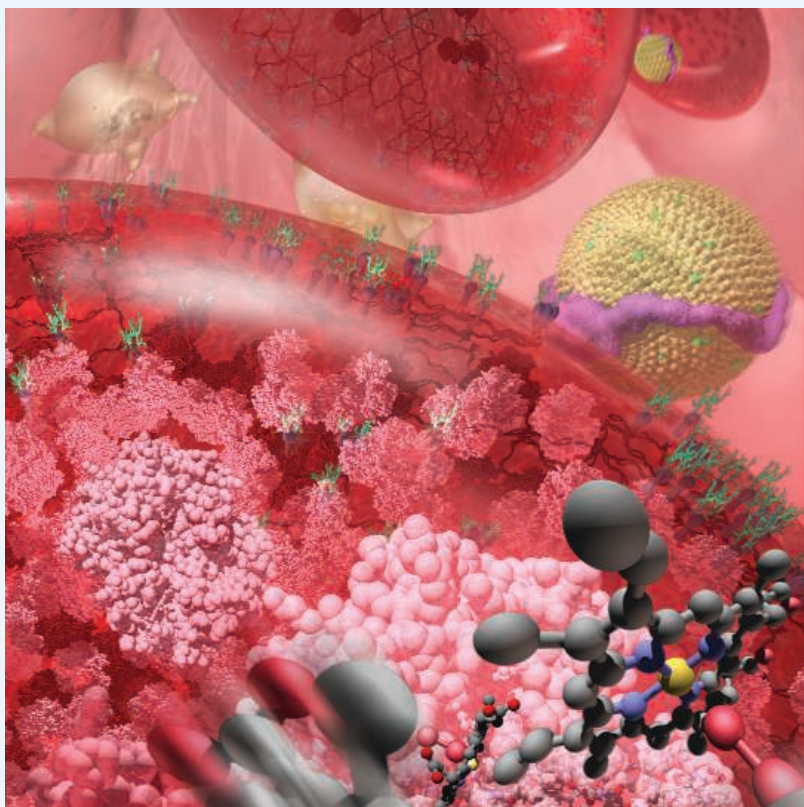
VISUALIZING
THE BIBLE

Chris Harrison, Carnegie Mellon
University, and Christoph
Römhild, North Elbian
Evangelical Lutheran Church

THE FIRST ILLUMINATED BIBLES were produced in the early Middle Ages by monks who painstakingly detailed illustrations for their sacred verse. Chris Harrison, a doctoral student at Carnegie Mellon University in Pittsburgh, Pennsylvania, and Christoph Römhild of the North Elbian Evangelical Lutheran Church in Hamburg, Germany, present an illustrated Bible with a modern twist. Römhild started with a list of verses in different versions of both the Old and New Testaments that referred to figures or ideas from earlier passages, then combed through both books for additional examples. Using a custom-built computer program, Harrison translated the trove of data into "Visualizing the Bible." Each bar on the graph along the bottom represents a chapter of the Bible; the bar length corresponds to the number of verses in the passage. The rainbowlike arcs represent references from a chapter in one book to a chapter in another. "It almost looks like one monolithic volume," Harrison says.
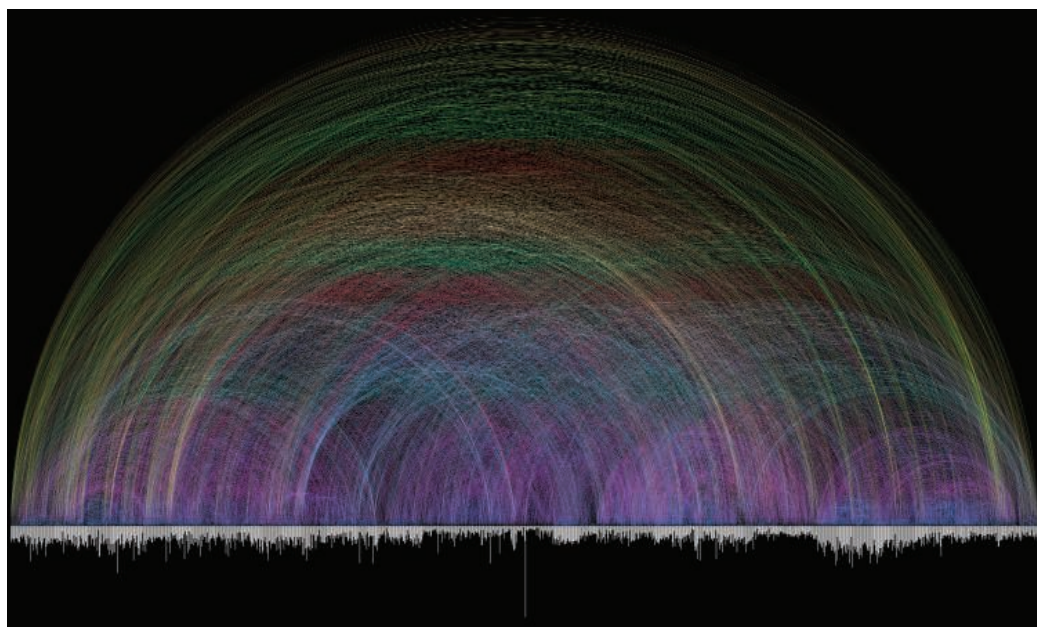
3D IMAGING OF MAMMALIAN CELLS WITH
ION-ABRASION SCANNING ELECTRON MICROSCOPY

Donald Bliss and Sriram Subramaniam, National Library of Medicine, NIH

THE DELICATE SWIRLS OF PINK AND GOLD IN THIS IMAGE COULD HAVE COME FROM BOTTICELLI'S brush, but there's nothing angelic about the subject, a melanoma cell. It is seen here by an ion-abrasion scanning electron microscope that uses a method of 3D imaging being developed at the U.S. National Cancer Institute. The microscope sends beams of gallium ions across an object, blasting away layers of the surface 20 nanometers at a time. By scanning each newly created surface, the microscope can compile three-dimensional images with unprecedented detail and resolution, says image creator Donald Bliss, a medical illustrator at the National Library of Medicine in Bethesda, Maryland. The images show almost too much detail—"It's like looking at a bowl of spaghetti suspended in clear Jell-O," he says—so Bliss chose to highlight some of the data. Here, he shows the nucleus as the dark sphere, engulfed by mitochondria (in pink) and endoplasmic reticulum (in gold).
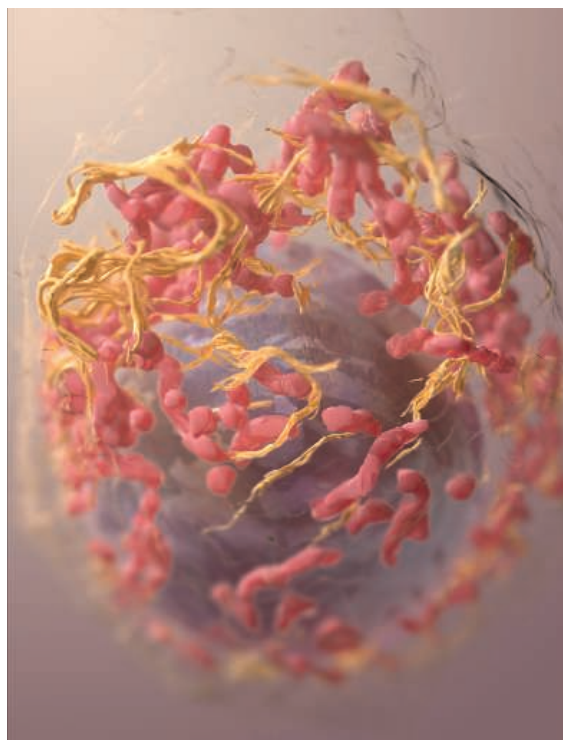
Informational Graphics

"MAD HATTER'S TEA" FROM ALICE'S ADVENTURES IN A MICROSCOPIC WONDERLAND

Colleen Champ and Dennis Kunkel, Concise Image Studios

WHILE WANDERING THROUGH THE FOREST OF WONDER-land, Alice stumbles upon three beetles having tea. That's not exactly how Lewis Carroll's classic tale goes, but this recreation of the Mad Hatter's tea could certainly belong in the story.

Freelance illustrator Colleen Champ produced her own version of the scene using micrographs by photomicrographer Dennis Kunkel. The goal was to demonstrate the fantastic nature of reality by arranging the actual images in fanciful ways, Champ says: "You cannot create anything yourself that hasn't already been created in nature."

She used Photoshop to transform three beetles into the Mad Hatter, March Hare, and the sleepy Dormouse. They sip tea at a table made of butterfly wings, set in a field of crystallized vitamin C while aphids fly overhead. A key beneath the main illustration identifies the source of each image, including the mold spores that make up the vast underground.

Kunkel plans to develop a series of children's books based on Champ's images.

The interplay between fact and fancy impressed the judges, who used the words "innovative" and "delightful" to describe the piece. Panel of finalist judges member Michael Keegan called it a "palatable introduction" to science, saying it provides an excellent way to attract children to the subject matter.

MAD HATTER'S TEA

The March Hare and the Hatter were having tea at it: a Dormouse was sitting between them, fast asleep, and the other two were using it as a cushion, resting their elbows on it, and talking over its head.

Death Watch Beetle · Monarch Butterfly Wing · Hatching Wigglers · Mold Fruiting Structure · Crystalized Vitamin C · Honeybee Stinger · Ladybug Beetle · Aphid Wasp · Mold Hyphae & Spores · Confused Flour Beetle

Stream micro-ecology: life in a biofilm

STREAM MICRO-ECOLOGY: LIFE IN A BIOFILM

Andrew Dopheide and Gillian Lewis, University of Auckland

ONE MAN'S SLIME IS ANOTHER MAN'S BIOFILM. RESEARCH TECHNICIAN Andrew Dopheide of the University of Auckland in New Zealand spends his days studying biofilm in streams under the direction of Gillian Lewis. Hoping to foster wider appreciation for his subject, Dopheide put together an informational graphic on the science of slime. He shows different magnifications of the primary biofilm dwellers: algae, bacteria, protozoa, cyanobacteria, fungi, and viruses. Factoids containing a brief description of the organism and its role in the system accompany the images. "There are all these quite fascinating things going on in this layer of slime. It's an important aquatic ecosystem," Dopheide says. The poster has already been distributed to schools and at scientific conferences in the country.

Interactive Media
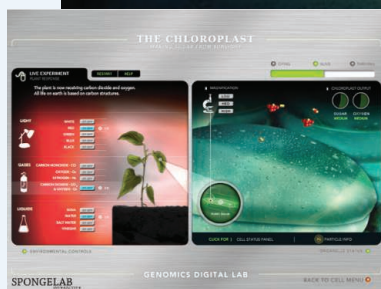
FIRST PLACE

GENOMICS DIGITAL LAB: PLANT CELLS

Jeremy Friedberg and Tommy Sors, Spongelab Interactive

"PLANTS ARE BORING." WHICH BIOLOGY teacher hasn't heard that complaint? Jeremy Friedberg and his colleagues at Spongelab Interactive in Toronto, Canada, along with Tommy Sors, a student at Purdue University in West Lafayette, Indiana, set out to give teachers an effective response with a computer program and educational game dedicated to plant biology. The Genomics Digital Lab uses flash animation and 3D graphics to present plant life in a dynamic light. As the name implies, the program focuses on the genomics approach to exploring biology. A few clicks of the mouse take users deep inside a plant cell, where they can choose among the chloroplast, mitochondria, and nucleus for further exploration. Each organelle lab contains a brief explanation of its function and a game in which students must pick the best light, water, and soil conditions for the plant to ensure the organelle's optimal performance. The goal is to help students understand the connection between the tiny organelles and the entire plant, Friedberg says: "We have to look at the whole and how something fits in that whole."

The Genomics Digital Lab enjoyed a surge of popularity when Apple Inc. posted the program on its Web site in January. To date, teachers in 22 different countries have downloaded the program, Friedberg says.

The interactive nature of the program earned high marks from the judges. "I remember studying very basic cell biology and being bored to death, but the fact that it was an interactive computer game you could get your hands on and see direct results of too much sun and not enough sun was very pertinent in this day and age when folks are so far removed from the plant and the planet," says panel of finalist judges member Malvina Martin.

HONORABLE MENTION

EXPLORING LIFE'S ORIGINS

Janet Iwasa, Massachusetts General Hospital
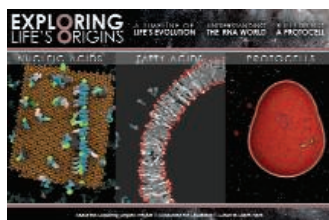
THE QUESTION OF HOW LIFE FIRST EMERGED lies at the heart of one of today's most contentious science debates. Biochemist Janet Iwasa wanted to fill an apparent gap in most documentaries on the origins of life. There were few visual explanations of how the first cells may have formed and operated on a molecular level, she says. So, while serving a fellowship at Massachusetts General Hospital, she produced this Web site using animation to illustrate topics such as how the original RNA polymers were assembled from nucleotides. See the Web site at www.exploringorigins.org.
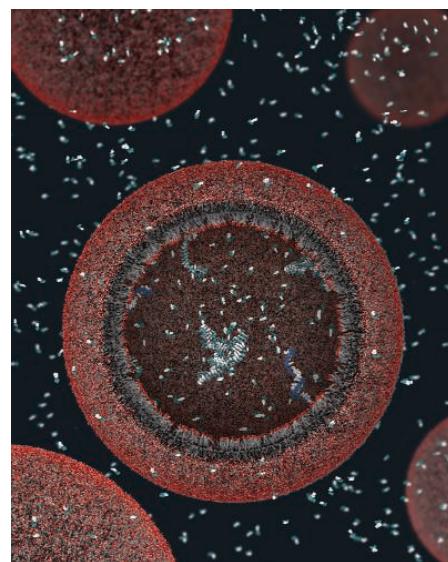
2008 VISUALIZATION CHALLENGE

Noninteractive Media

A WINDOW INTO LIFE

Travis Vermilye and Kenneth Eward

EVEN WHILE AT REST, OUR BODIES PULSE WITH furious activity. Neurons fire, cells divide, and proteins form, only to be dismantled in short order. This movie shows vignettes of the microscopic plane of life on which our everyday lives depend. Designed for display within the Cincinnati Children's Hospital Medical Center, the movie explores some of the basic science behind the hospital's research projects. Freelance illustrator Kenneth Eward and freelance animator Travis Vermilye, who collaborated to produce the film, give a whirlwind tour of the assembly-line process by which RNA builds proteins. They also sneak up close to a neural synapse as it fires a message to a nearby muscle fiber and show how the eye develops from the embryonic stage to the mature form. The goal is to present the dynamic complexity of life on the smallest scale: "I'd like people to be inspired by the beauty that goes on inside of us," Eward says.
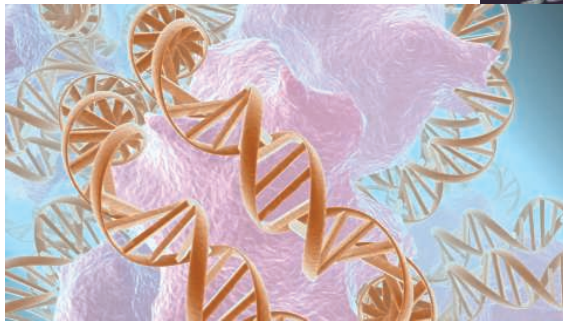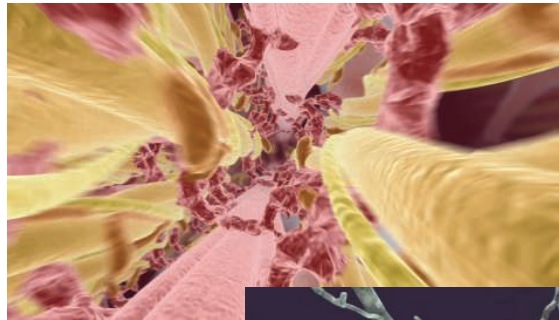
SMARTER THAN THE WORM

Mirjam Kaplow and Katharina Strohmeier, Fraunhofer FIRST

WHEN THE DREADED ERROR MESSAGE FLASHES on the screen, it's easy to envision an army of malevolent gremlins wreaking havoc on your computer. The real mechanism of a computer worm or virus isn't quite that dramatic, but producer Mirjam Kaplow and Katharina Strohmeier of Fraunhofer FIRST in Berlin, Germany, play on that tension to explain how those pests operate and how computer software protects against them. "We came up with the idea of making a movie with the symbol on a metaphorical level," Kaplow says. The story takes place at the gates of a fortress city, where a guard examines each visitor before granting them access. But simple disguises—a new bow tie or a pair of sunglasses—confuse the guard, and he lets a worm slip through. While the city burns, the narrator explains how a new type of software can keep a computer smarter than the worm, whatever that worm looks like.
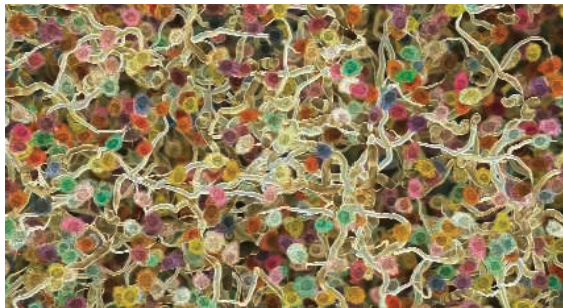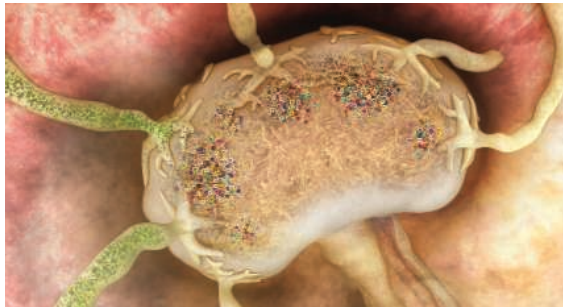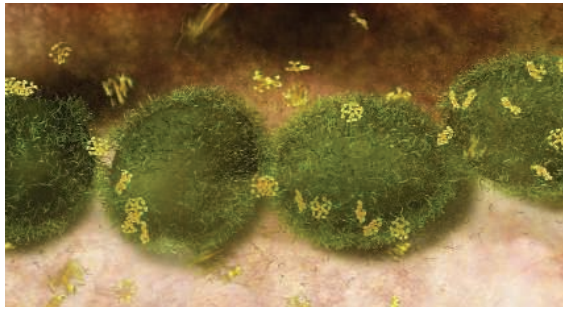
FIGHTING INFECTION BY CLONAL SELECTION

Etsuko Uno and Drew Berry, The Walter and Eliza Hall Institute of Medical Research

IN 1960, AUSTRALIAN IMMUNOLOGIST FRANK BURNET won a Nobel Prize for his contributions to immunology. Etsuko Uno and colleagues at the Walter and Eliza Hall Institute of Medical Research in Melbourne, Australia, explain Burnet's clonal selection theory in an animation of the body's response to *Streptococcus pyogenes*, the bacterium that causes strep throat. Proteins from the invader enter the lymph node and grab the attention of one of billions of B cells. That B cell then clones itself thousands of times and sends antibodies via the bloodstream to the infection site. There, the antibodies bind to the strep bacteria, acting as a red flag that alerts other immune system cells to destroy the infectious agent. "We hope that the animation will pique people's interest in how the immune system works and that they will appreciate the impact of Burnet's clonal selection theory on our understanding of the immune system," Uno says.

2008 VISUALIZATION CHALLENGE

COMMENTARY

The burdens
of atlases

1780

Slipping secrets down
noisy channels

1783

LETTERS I BOOKS I POLICY FORUM I EDUCATION FORUM I PERSPECTIVES

LETTERS

edited by Jennifer Sills

Fixing the Leaky Faucet

A. I. LESHNER'S EDITORIAL "JUST GIVE THEM GRANTS" (16 May, p. 849) is an urgent call for dedicated funding for new investigators in science. However, without a means to sustain new investigators once their laboratories have been established, another crisis will quickly follow: the inability to retain the talent brought to the bench. This leaky-faucet phenomenon is already well known to women in medicine and science, with much good will to slow down these departures but little resolution in sight. The academic and funding community must be committed to the full length of the science career, not just the early part of it. Why recruit if we cannot retain? To do so will only create disillusionment and distrust among those in whose hands the future of science lies.

JUDY ILLES

Department of Neurology, National Core for Neuroethics, The University of British Columbia, Vancouver, BC V6T 2B5, Canada. E-mail: jilles@interchange.ubc.ca

the backs of senior investigators who, despite clearly being deadwood, are hogging multiple grants. But then I wondered what will happen to me in the phase between being a new investigator and a senior investigator. Surely there must be some intermediate step. The transformation from plucky young innovator to conservative graybeard cannot be instantaneous. After that first RO1, I will have some publications and some data, but not as many as my more senior competitors. And the funding situation for me will be even tighter than before, because a significant fraction of funds will be going to those undeserving, knee-biting, new investigators. Additionally, I will have reached a stage where I have significant responsibilities—graduate students and postdocs will be depending on me for their career advancement and livelihood. So I am left asking, what will happen to new investigators once their honeymoon is over?

SANJAY S. P. MAGAVI

Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. E-mail: smagavi@mit.edu

Redefining Academic Success

A. I. LESHNER MAKES SUCH A COMPELLINGLY simple recommendation in his 16 May Editorial ("Just give them grants," p. 849) that one cannot but wonder why it need be made at all. If junior academic scientists need a government-funded grant to launch their independent research career, why not just give them grants? Problem solved. However, if the academic research community is really going to tackle what is, despite the Editorial's straightforward prose, a very complicated issue, then it should also consider another seemingly simple question extracted from the Editorial's first sentence. Why is it that securing external funding for independent research is a "gold standard" for academic success, particularly in the first few years of a career spanning decades? Shouldn't the early investment in a junior faculty member's scholarly research be the responsibility of the institution hiring him or her? Might not considerations of success also include the originality of the individual's

research, the contributions the research could make to the intellectual content of his or her chosen field of research, and the value of the individual as a colleague? Surely there are ways for institutions to develop internal metrics of success. So, here is another simple recommendation: It is time for academic institutions to stop ceding their promotion and tenure decisions to the NIH and other external funding bodies.

SUSAN M. FITZPATRICK* AND JOHN T. BRUER

James S. McDonnell Foundation, St. Louis, MO 63117, USA.

*To whom correspondence should be addressed. E-mail: susan@jsmf.org

Caught in the Middle?

AS A POSTDOCTORAL FELLOW ENTERING THE market for biology faculty positions, I was happy to hear that 25% of NIH Research grants are going to new investigators who have never received an RO1 (Editorial, "Just give them grants," A. I. Leshner, 16 May, p. 849). My sense of schadenfreude was fulfilled to hear that these funds will come off

Just Give Them Fellowships

IN THE RECENT EDITORIAL HIGHLIGHTING THE issues faced by young academics in securing funding for their own research ("Just give them grants," 16 May, p. 849), A. I. Leshner touches on an important point: the subversion of personal research interests during postdoctoral training periods. As an example, UK government-funded research

CREDIT: JUPITER IMAGES
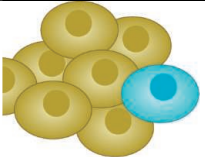
How cancer spreads

1785

Bidding high!

1788

councils typically expect grant recipients to be appointed at a higher education institute, with a minimum position of Lecturer, before applying for a research grant. As a result, the pressure on freshly minted Ph.D.'s in academia is, as stated, to follow the path of postdoctoral research on established projects, rather than trying to secure their own funding.

Clearly, the UK research councils place strong emphasis on the training of postgraduates, but there appears to be little incentive for those students to remain within academia in the hope of pursuing their own lines of research by obtaining individual postdoctoral fellowships (*1*). Such fellowships provide opportunities for young scientists to "make their mark" in their respective fields without being tied to lines of research that they do not wish to pursue. By awarding more fellowships, funding organizations may retain more individuals to contribute to the continuity of scientific enterprise and, in turn, fellows may find getting that first grant or tenure position a little bit easier.

ANGELO P. PERNETTA

CEH Wallingford, Centre for Ecology and Hydrology, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB, UK. E mail: anpp@ceh.ac.uk

Reference
1. Research Councils UK (www.rcuk.ac.uk).

Destabilizing the Pyramid Scheme

A. I. LESHNER'S EDITORIAL "JUST GIVE THEM grants" (16 May, p. 849) suggests yet another well-meaning "fix" for the poor funding support for advancing research careers of post-docs and young investigators in the United States, especially in the biomedical fields. Setting "funding quotas" by earmarking a percentage of new grants, those from the NIH in particular, to investigators younger than a certain age probably will not aid a situation that is, largely, a pyramid scheme.

Pyramid schemes provide considerable incentives for those at the top [funded principal investigators (PIs), tenured faculty, and most medical school faculty] and virtually no tangible incentive to those at the bottom, who support the scheme at the laboratory bench (graduate students, postdocs, and research associates). Perhaps it is time for some disincentives for those at the top. Some suggestions follow.

(i) Any grant proposal that gives salary support for postdocs must also include funds for postdoc-only projects (mini-grants within a grant). (ii) Any PI who proposes to put postdocs on the grant payroll should have documentation that he or she has also done a stint as a postdoc. Who better knows the value of mentoring than those who have been mentored? (iii) Postdoc

training is a disheveled cottage industry. Establish a central clearinghouse of "post-doc specialists" akin to "Matching Day" for medical school graduates seeking advanced training in limited residency training positions. (iv) If postdocs are to be a necessary part of the research enterprise, then PIs, or their departments or institutions, should provide some guarantee of financial support beyond the tenure of a particular grant to those postdocs who provide credible service to that grant but who cannot find their own support elsewhere.

DUAINE R. JACKOLA

8436 Second Avenue South, Bloomington, MN 55420, USA. E-mail: drjackola@q.com

Biotechnology Innovation in Africa

AFRICA IS PRESENTLY AT THE PRECIPICE OF A socioeconomic renaissance. However, diseases such as malaria, AIDS, and hypertension remain common and important health problems facing the continent. The recent Policy Forum by T. J. Tucker and M. W. Makgoba ("Public-private partnerships and scientific imperialism," 23 May, p. 1016) should invoke further discussions on new approaches for increasing the effectiveness of global efforts against neglected African diseases.

In the 1970s, 70% of resource flows from the United States to the developing world were from official development assistance and 30% were private. Today, 85% of resource flows from the United States to the developing world are private and 15% are public. These changes in resource flows reflect the emergence of the private for-profit sector and the nongovernmental sector as crucial participants in the development process (1). They have formed many new alliances and programs in addition to government aid. Unfortunately, when funds for these programs run out, the progress often stagnates or even reverses. Few public and private donor programs exist to support more sustainable programs, such as small indigenous African bioscience businesses that are evolving biotechnological innovations specifically relevant to the region.

Developing local biotechnology capacity is essential for ensuring availability and access of health care products in a sustainable manner. Several governments in sub-Saharan Africa (such as Nigeria and South Africa) recognize this and have increasing public sector support for biotechnology innovation and entrepreneurship to encourage small indigenous biotechnology companies that are working to translate relevant research discoveries to usable products. National and regional public policies and priorities are encouraging the local development and manufacture of essential rapid diagnostics and genuine medicines that are critical to health care needs of the people, as a way of making these products and services more readily accessible to more people. In response, an increasing number of entrepreneurial scientists of African descent (led by Africans in The Diaspora) are establishing local, small, socially responsible biotechnology enterprises. These efforts are inspired primarily by necessity and a focus on translating relevant discoveries to products and services that address regionally prevalent diseases.

A model that has not gained broad acceptability among private donors is direct support in the form of pass-through grants to small indigenous for-profit bioscience businesses. Robert Grant had proposed a similar context in his "Research in situ" model (2, 3). By working with indigenous for-profit bioscience companies, multilateral funding organizations and agencies can potentially deliver more sustainable change. This is especially crucial because many developed nations have modeled small businesses as the core of their biotechnology development strategy, strengthened through government and investor-backed small business grants and loan programs. Streamlined donor support to indigenous small bioscience businesses can enable the development of specific new products and services consistent with the socio-economic needs of the continent. Additionally, through expanding collaborations with universities and institutes, the indigenous biotechnology firms are evolving to create open avenues of knowledge sharing to create these products in a sustainable manner. This can potentially drive the development of biotechnology on the continent.

EDDY C. AGBO,[1]* SIMON AGWALE,[2] CAMELLUS O. EZEUGWU,[3] BOITUMELO SEMETE,[4] HULDA SWAI,[4] ANTHONY IKEME,[5] RICHARD I. SOMIARI[6]

[1]Fyodor Biotechnologies, 26 Ogui Road, Enugu, Nigeria, and 3607 Frankford Avenue, Baltimore, MD 21214, USA. [2]Innovative Biotech Ltd., 1 Abdu Abubakar Street, GRA, Post Office Box 30, Keffi, Nasarawa State, Nigeria. [3]The Johns Hopkins University School of Medicine, Department of Medicine (JH Cardiovascular Group Inc.), Baltimore, MD 21201, USA. [4]Council for Industrial Scientific Research (CSIR)—Polymer and Bioceramics, Post Office Box 395, Pretoria, South Africa. [5]Clintriad Pharma Services, Exton, PA 19341, USA. [6]ITSI-Biosciences, Johnstown, PA 15904, USA.

*To whom correspondence should be addressed. E-mail: eddy.agbo@fyodorbio.com

References

1. USAID Global Partnerships (www.usaid.gov/our_work/global_partnerships/gda).
2. R. M. Grant, *Nat. Methods* **4**, 887 (2007).
3. Editorial, *Nat. Methods* **4**, 877 (2007).

CORRECTIONS AND CLARIFICATIONS

Reports: "Alignment uncertainty and genomic analysis" by K. M. Wong *et al.* (25 January, p. 473). C. Dewey, A. Schwartz, N. Bray, and L. Pachter kindly directed our attention to an inconsistency in Fig. 1, which shows six different estimated trees for seven different alignments of the open reading frame (ORF) YPL077C, and the Supporting Online Material containing the maximum likelihood estimates for the 1502 ORFs that we examined. When equally likely trees are accounted for, maximum likelihood yields only four different trees for YPL077C. We intended to illustrate an extreme example in which alignment uncertainty produces different estimates of phylogeny, and not to select among equally likely trees to make the differences as great as possible. Indeed, there was no reason to do so, because we could have illustrated the point with five other ORFs, all with one estimated tree for each alignment and resulting in six different trees for the seven alignment treatments (see the Supporting Online Material). Of potentially more importance, however, our results did not account for equally likely trees, something that occurs in 1.5% of the phylogenetic analyses. Figure 1 repeats the analyses performed in the original Report and accounts for equally likely trees. As before (Fig. 2A), we see a significant positive correlation between alignment distances among alignment treatments and the distances between trees estimated from the alignments. Accounting for equally likely trees does not change the relation between alignment variability and phylogeny estimation we originally discussed.

Fig. 1. Positive correlation between the Robinson and Foulds [D. Robinson, L. Foulds, *Math. Biosci.* **53**, 131 (1981)] measure of topological distance among trees estimated from different alignment methods and alignment variability among alignment treatments (Spearman's rank correlation: $r_s = 0.52$, $P < 0.0001$; note that the correlation coefficient changes from $r_s = 0.53$ to $r_s = 0.52$ when equally likely trees are accounted for).
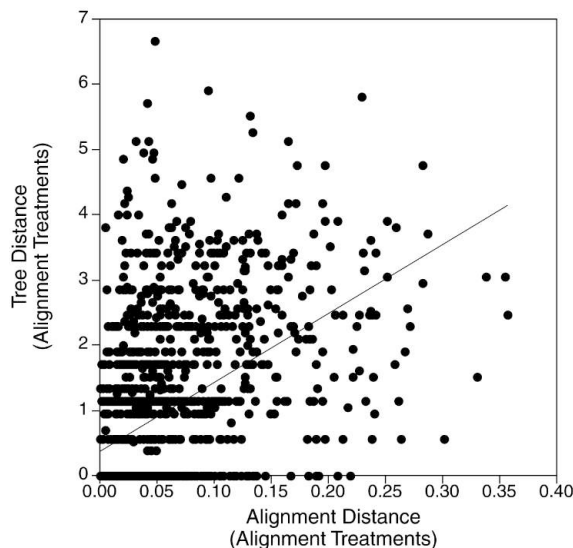
ECONOMICS

For Equality, Education Matters

Thomas Lemieux

Throughout most of the 20th century, the economic performance of the United States was simply remarkable. Fueled by technological advances and an ever-more-educated workforce, productivity grew steadily and standards of living of each generation far exceeded those of the preceding one. Defying the old adage that there is a tradeoff between equality and efficiency, all this growth was achieved while the gap between rich and poor was declining. Furthermore, America's education system delivered on the promise of equal opportunity by making a good education, the best gateway to well-paying jobs, accessible to most irrespective of their background and circumstances. By the 1970s, the United States was not only the richest country in the world, it also boasted the most educated population while inequality was no higher than in most other rich countries.

But then something happened. Beginning in the 1970s, productivity stalled and inequality started growing rapidly. The combination led to declining incomes and standards of living for a substantial fraction of the U.S. population. To make things worse, the great progress in higher education that had run through most of the 20th century came to an abrupt halt. A college degree would have guaranteed to most young people standards of living as good as or better than those of their parents, but the fraction of young Americans completing college hardly changed in the last quarter of the 20th century.

Not surprisingly, these developments have had many economists scratching their heads and wondering what went wrong in the last few decades. With inequality now reaching highs not seen since the Great Depression, considerable effort has gone into understanding why inequality increased so much. Many have pointed their finger at technological change. For sure, technological progress has always been a leading source of the growth of nations. But the benefits of technological change are not evenly shared among the whole workforce (1). Globalization and off-shoring, institutional factors like the minimum wage and unionization, and broader social norms about how much inequality is acceptable in a society have also been suggested for explaining the growing gap between rich and poor.

The key contribution of Claudia Goldin and Lawrence Katz's masterful *The Race Between Education and Technology* to the inequality debate is to take another look at the recent changes through the lenses of history. Doing so yields a number of important insights. In particular, history tells us that though technological change may be a source of growing inequality, this cannot be the whole story. Indeed, there have been other historical episodes of fast technological change, such as electrification in the 1910s and 1920s, that did not result in higher inequality. Why is that? The book's main conclusion is that in those days rapid technological change was accompanied by stunning gains in educational achievement that provided enough qualified workers to meet the demands of an increasingly technologically sophisticated economy.

As is often the case in economics, this story boils down to supply and demand. Technological change increases the demand for highly educated workers who can most effectively use the new technologies in the workplace. If the supply of highly educated workers does not increase fast enough, however, firms start bidding up the wages of these workers to attract them, while others who lack proper qualifications lose ground. So, as the book's title suggests, whether inequality increases or not is best thought of as an ongoing race between education and technology.

Combining this simple but appealing idea with a deep knowledge of the histories of the U.S. labor market and educational institutions, Goldin and Katz (economics professors at Harvard) conclude that whereas education was winning the race for most of the 20th century, technology caught up in the 1970s and has since prevailed. The authors' most insightful point is that the root cause of the recent growth in inequality is not faster technological progress during the past three decades but rather the surprising stagnation in the level of education of young Americans.
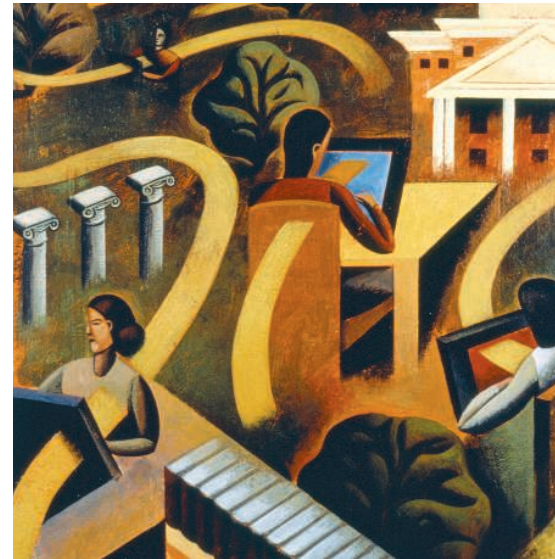
Besides addressing the increasing inequality, the book provides a fascinating history of the American education system over the last century that is a must for anybody interested in this important topic. Although the United States may be best known for its top private universities that are the envy of the world, Goldin and Katz convincingly show that the source of the country's long educational supremacy had much more modest origins. By 1940, most young Americans attended publicly funded high schools, a situation simply unthinkable in the rest of the world at the time. Post–World War II, the G.I. Bill continued from where the high school movement of 1910 to 1940 had left off, opening college to the masses.

But things have very much changed since then, and many other countries have now caught up. Young Americans no longer have the educational advantage that their parents or grandparents enjoyed over the rest of the world. For the United States to regain its leadership role and win the race against technology, more resources must go to making sure high school students are college-ready and that those who are college-ready have financial access to higher education. The difficulty is that doing so will cost money. As the analysis presented by Goldin and Katz indicates, the well-being of the new generations will critically depend on whether America is up to this challenge.

The reviewer is at the Department of Economics, University of British Columbia, Vancouver, BC V6T 1Z1, Canada. E-mail: tlemieux@interchange.ubc.ca

References
1. D. Autor, L. Katz, A. Krueger, *Q. J. Econ.* **113**, 1169 (1998).

CREDIT: CURTIS PARKER/GETTY IMAGES

HISTORY OF SCIENCE

Regime Change in Scientific Depiction

Alan Richardson

It is nearly universally accepted that science aims for an objective view of the world—and that this is a virtue of science. Indeed, it can seem axiomatic that if you are seeking the truth about nature, then you ought to approach nature objectively. What other option is there? A knowledge-seeker, surely, ought not approach nature with a socially inculcated or individual bias, with emotional attachment to a theoretical view already in hand, or with an aesthetic or moral or financial interest in the subject. Therefore, when scientists hear that Lorraine Daston and Peter Galison, two of the world's most distinguished historians of science, claim in *Objectivity* that, in a crucial sense, science only began to strive for objectivity in the 19th century and is now moving away from it, they might worry that once again social historians of science are attempting to debunk the practices and goals of science.

This would be an unfortunate reaction, for although Daston (Max Planck Institute for the History of Science, Berlin) and Galison (Harvard University) are indeed historicizing the notion of objectivity and making important claims about its place in the history of science, theirs is not a project of unmasking or debunking. The point of their enterprise is neither to look at how claims to objectivity disguised powerful social or gender bias nor to dismantle the concept of objectivity. They take for granted the complexity in the notion of objectivity, as they take for granted both the social character of scientific practice and the complicated and varied social places of science across times and cultures. What they are interested in are the 19th-century rise of an account of science that stressed science's objectivity and how the science framed by that account differed from science earlier and later.

Objectivity is, for Daston and Galison, a specific epistemic ideal of science, entailing certain epistemic virtues the scientist ought to have. These virtues are motivated by specific anxieties about obstacles to knowledge—and they are secured not only by exhortations in their favor but also by practices that structured scientific work in accordance with them. Those practices were forms of "collective empiricism" employed to implement the regulative ideal of objectivity. None of this means that other respectable and equally empirical regimes of science are not possible, and the authors argue that such regimes have been and are now being pursued.

The authors provide their argument not in abstracto but through examination of distinct practices of one widespread type of scientific activity: the making of scientific atlases. They

Observer as machine. In his figures, such as these blood crystals, Otto Funke "attempted to reproduce the natural object in its minutest details … above all things prohibiting the slightest idealization."

contrast regimes of objectivity (mechanical and structural) with truth-to-nature and expert judgment. At the end, they suggest that recent image galleries in science are again moving in new directions: away from representation of already existing objects and toward presentation of objects made in the very process of depicting them—as in the famous nanotechnological depiction of xenon atoms formed into an advertisement for IBM, an image produced by the very probe that arranged the atoms.

The authors' argument here is complicated but fascinating (and, because the argument is about images, the book is beautiful). Daston and Galison hold that notions of objectivity and subjectivity as they have been deployed since the 19th century were forged in the crucible of German philosophy and expressed new anxieties about the self as the chief obstacle to accurate knowledge of the world. The 19th-century atlas makers, therefore, carefully attempted to remove their own judgment from their images, often deploying new photographic technologies to keep the images free from their own interventions. Thus, Daston and Galison call the scientific project of the day "mechanical objectivity," with the machine as the mindless, will-less helper (the camera renders depiction objective precisely because it neither sees nor interprets). The authors contrast this "blind sight" with an 18th-century practice of explicit intervention into atlas images, images meant to depict not individual specimens but rather the ideal type of the object. These earlier images required the theoretical knowledge of the atlas makers, who carefully policed their illustrators' drawings not to erase idealizations but to guide them. Similarly, Daston and Galison argue that by the mid-20th century a more self-confident and widely dispersed community of scientific experts again changed the practices of making and reading atlases, substituting a regime of expert judgment for the practices of mechanical objectivity. These later scientists again often substituted drawings for photographs and again self-consciously altered images, but now not to depict ideal types but rather to help train the reader into having an expert eye.

Are not all these regimes, because they are all concerned with accurate representation of nature, in some sense, regimes of objectivity? Certainly. But to rest content with this "in some sense" would be to limit, within historical inquiry, the motive to treat objects of inquiry with the specificity and precision that guide Daston and Galison's "mechanical objectivity." It would also leave us fewer conceptual tools for thinking about both the newly aestheticized, newly commercialized, newly interactive, and newly object-creating images of 21st-century science and the new type of scientist who makes such images.

The reviewer is at the Department of Philosophy, University of British Columbia, Vancouver, BC V6T 1Z1, Canada. E-mail: alanr@interchange.ubc.ca

References

1. O. Funke, *Atlas of Physiological Chemistry* (Cavendish Society, London, 1853).

10.1126/science.1162742

CREDIT: OTTO FUNKE, PLATE 10, FIGURE 2, IN (1)

ASSESSMENT

School Performance Will Fail to Meet Legislated Benchmarks

Federally mandated progress goals may translate into widespread failure of California elementary schools.

M. J. Bryant,[1,4] K. A. Hammond,[1] K. M. Bocian,[2] M. F. Rettig,[3] C. A. Miller,[1] R. A. Cardullo[1]*

There is widespread concern that U.S. students entering college do not have the educational foundation to succeed in introductory science courses. Mathematics is of particular interest because proficiency in mathematics may predict performance in introductory college science courses (1). In the United States, numerous educational reforms (2) focus on the accountability of schools for proficiency (mainly in mathematics and English Language Arts). However, agreement on proficiency goals that accurately reflect the needs of a diverse student population has been problematic.

Downward trends in U.S. student test performance from 1963 to 1980 led to recommendations for increased accountability for schools and school districts (3). The Improving America's Schools Act (IASA), Title I, legislation in 1994 (4) mandated that states should set standards by 1999 and have in place assessments to measure achievement by 2000. The "No Child Left Behind Act of 2001" (NCLB), which mandates that 100% of students will be proficient in English Language Arts (ELA) and mathematics by 2014, is the strongest current implementation of this accountability (5).

Under NCLB, a school's progress toward having 100% of students proficient is monitored by Adequate Yearly Progress (AYP), a series of calculated academic performance factors for each state, local education agency, school, and numerically significant student subgroup within a school [English language learner (ELL), socioeconomic status, etc. (table S1)]. All states must use as AYP criteria: (i) the percentage of students scoring proficient or advanced on state standards tests in ELA and mathematics, (ii) standardized testing participation rate, and (iii) increases in the high school

[1]Department of Biology, University of California, Riverside, Riverside, CA 92521, USA. [2]The ALPHA Center, University of California, Riverside, Riverside, CA 92521, USA. [3]Department of Chemistry, University of California, Riverside, Riverside, CA 92521, USA. [4]School of Critical Studies, California Institute of the Arts, Valencia, CA 91355, USA.

*Author for correspondence. E-mail: cardullo@ucr.edu

graduation rate. However, each state determines its own specific assessments and cutoff scores for proficiency, and these vary widely from state to state (6). All states are required to report proficiency in mathematics and ELA.
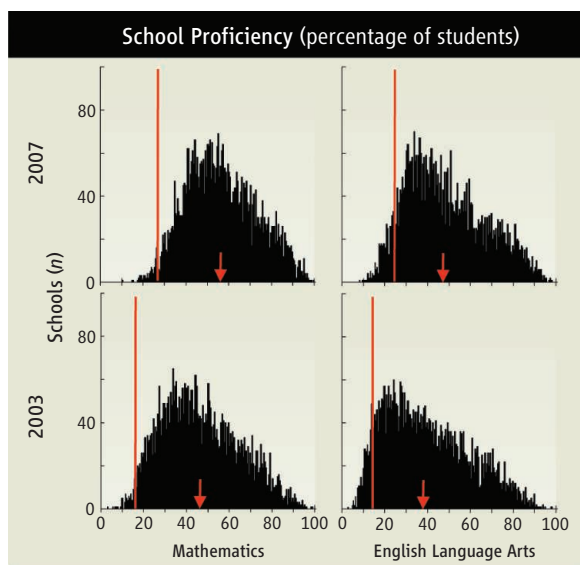
Divergent proficiencies of California schools. The proficiency of California elementary schools in 2003 (n =4850) and 2007 (n = 4917) is defined by the percentage of total students in that school who achieved a score of proficient or advanced. Red line, AYP benchmarks; arrow, statewide mean school proficiency.

California also includes additional subjects such as science as part of its Academic Performance Index (table S1).

Although NCLB mandates an evidence-based, research-driven approach to educational reform, each state determines how the 2014 benchmark of 100% proficiency will be achieved. This presumably allows each state to look at local conditions and to tailor programs, assessments, and sanctions to reach the 2014 target. In reality, lack of consensus about proficiency benchmarks makes it difficult to assess the value of NCLB as a research-driven, national-scale force for education reform (7–13). Nonetheless, it is possible to analyze the data from individual states to determine whether NCLB targets will be met in that state (13).

California, with over one-eighth of the U.S. population, has comprehensive data for AYP analyses. The state has (i) a large number of

schools and accessible test results, (ii) a sufficient number of population subgroups identified by NCLB legislation, and (iii) a consistent standardized assessment tool for monitoring gains in student learning across schools. We used data from California elementary schools to ask if achievement gains will reach 100% proficiency by 2014. Although this question has been addressed elsewhere (14, 15), the analyses reported here raise questions about the distribution of school-level achievement in terms of the specific test subject and vulnerable subgroups of students.

Evaluating School Performance

We collected data available through the California Department of Education's accountability progress reporting system and extracted the following data for all students and numerically significant subgroups within elementary schools: the number of students tested and the percentage of students scoring proficient or advanced for both the mathematics and ELA test. We employed linear, polynomial, and logistic models to make comparisons with state and federal growth targets for students scoring proficient or advanced within schools (16).

A school is held accountable to a specified percentage of students scoring proficient or advanced and, if it fails to meet this, the school may be sanctioned. This holds true if either the targeted percentage of all students in a school or the percentage of any numerically significant subgroup fails to reach the annual benchmark for either mathematics or ELA tests. We define the AYP criterion that matters from a legislative and school perspective as the lowest percentage of students scoring proficient or advanced on either test and from any group or subgroup (SOM), because it is the lowest percentage that may initiate sanctions.

Insufficient Achievement Gains

Students in California elementary schools show a wide range of proficiencies in mathematics and ELA (see chart, above). From 2003 to 2007, three evident trends are that (i)

overall performance is higher for mathematics than for ELA; (ii) in both subjects, the mean school proficiency increased; and (iii) there has been negligible reduction in interschool variation as would be implicit in a target of 100% proficiency.

For ELA, the logistic model best reflected the observed achievement gains. For mathematics, it was the polynomial model (table S2). All models estimated average annual growth rates under 4% from 2003 to 2007. However, to meet the goal of 100% proficiency mandated for 2014, the AYP targets designated by Californian and federal legislation will increase annually by 10.8% (ELA) and 10.5% (mathematics) from 2008 onward. We predict that ~50% of all California elementary schools will fail to meet AYP by 2011 (fig. S1). Our candidate models permitted optimistically rapid acceleration in student achievement. However, both logistic and polynomial models indicate a deceleration in achievement gains. The data currently available suggest that a large number of schools (and thus students) will not make the 2014 target of 100% proficiency.

We asked which group or subgroup reflected the lowest level of proficiency for each school and if that level of proficiency was based on the mathematics or the ELA test (see chart, above). For most schools, the greatest risk of failing AYP lies with ELA proficiency. Additionally, it is the Socioeconomic Disadvantaged and ELL subgroups that are most likely to fail to meet AYP in California. These data suggest that very little change in this pattern can be anticipated over the next 5 years (fig. S2).

Modest Gains Will Not Meet Goals

We demonstrate that nearly all elementary schools in California will fail to meet the AYP requirements for proficiency by 2014. Recent gains in the statewide percentage of students scoring proficient or advanced in California have been used to suggest that accountability initiatives such as NCLB are effective in causing increased student learning. However, these documented gains for state-level performance conceal the fact that it is individual schools that are sanctioned under NCLB (17). Ignoring the variation around statewide averages obscures important patterns and information that may guide future professional development, the cre-

ation of new assessment tools, or legislation.

California's large, diverse public school population allows data on trends in various subgroups to be examined for significant differences with regard to AYP. Within a school, the performance of just one subgroup on one test subject can determine whether that school will meet AYP and escape sanctions. Redirecting the focus to the lowest performing subgroups at each school results in a different distribution in school performance that reveals a profound acceleration in the date at
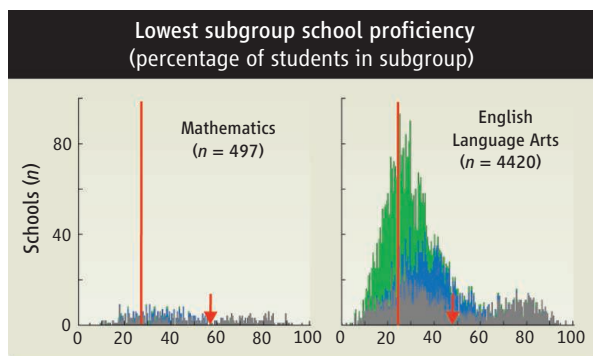
Susceptible subgroups and subjects in California schools. In California elementary schools (n = 4917), more school proficiency failures are due to ELA within the ELL student subgroups (year 2007). The subgroup school proficiency is defined by the lowest percentage of students within a subgroup achieving a score of proficient or advanced. ELL students, green; Socioeconomic Disadvantaged students, blue; all other students, gray. Red line, AYP benchmarks; arrow, statewide mean school proficiency based on all students.

which schools will fail to meet AYP.

The weakness of ELA progress suggests that more emphasis should be placed on ELA, although that would be a shortsighted solution. Schools are also in need of support in mathematics, for which we predict nearly 100% failure, with current approaches, to meet AYP by 2014. A student's success in further education depends on the ability of that student's schools from kindergarten through completion of high school to deliver an effective curriculum that includes both ELA and mathematics, as well as the social and natural sciences. The data examined here suggest that the elementary schools in California are not meeting these teaching challenges and that resources to improve the situation are not available or have not been applied to date.

Accountability of schools is useful only to the degree that the assessments are valid and reliable. The assessment data must also be analyzed in a meaningful way that can positively affect instruction. By focusing the attention on the central tendencies of the highest performing students (proficient and advanced), we risk ignorance of the progress of the lower-performing students—ignorance that potentially leaves behind those who were to be served by NCLB.

The richness in the current California data set is compromised because the categories below proficient are not routinely examined for gains. (California has three lower-performance levels: basic, below basic, and far below basic). Because the lower proficiency groups are not usually analyzed, we cannot adjust and apply interventions for these troubled students.

Future educational reform legislation would benefit from a comprehensive research focus. Reforms should tie educational experiences to instructional challenges experienced by a particular school, focusing each school's own resources to serve its own unique student population. As called for in "A Nation at Risk" (3), we must not lose sight of the importance of educating our children well. From providing a well-educated voting public to improving the numbers, preparation, and diversity of our college-bound students, this mission deserves our greatest effort.

References and Notes

1. P. M. Sadler, R. H. Tai, *Science* **317**, 457 (2007).
2. P. L. Kimmelman, *Implementing NCLB: Creating a Knowledge Framework to Support School Improvement* (Corwin Press, Thousand Oaks, CA, 2006).
3. National Commission on Excellence in Education, *A Nation at Risk: The Imperative for Educational Reform* (U.S. Government Printing Office, Washington, DC, 2003).
4. U.S. Public Law 103-382 (1994).
5. *No Child Left Behind Act of 2001*, U.S. Public Law 107-110 (2002).
6. J. Cronin *et al.*, *The Proficiency Illusion* (The Thomas B. Fordham Institute, Washington, DC, 2007); www.edexcellence.net/detail/news.cfm?news_id=376.
7. R. L. Linn *et al.*, *Educ. Res.* **31**(6), 3 (2002).
8. J. Powers, *Educ. Pol.* **17**, 558 (2003).
9. M. Russell *et al.*, *Teachers College Rec.* **106**, 2102 (2004).
10. J. Kim, G. Sunderman, *Educ. Res.* **34**(8), 3 (2005).
11. L. B. Resnick, *Educ. Meas. Issues Practice* **25**, 33 (2006).
12. B. Fuller *et al.*, *Educ. Res.* **36**(5), 269 (2007).
13. Center on Education Policy, *Answering the Question That Matters Most: Has Student Achievement Increased Since No Child Left Behind?* (Center on Education Policy, Washington, DC, 2007); www.cep-dc.org/index.cfm?fuseaction=document.showDocumentByID&nodeID=1&DocumentID=200.
14. N. Chudowsky, V. Chudowsky, *Many States Have Taken a "Backloaded" Approach to No Child Left Behind Goal of All Students Scoring "Proficient"* (Center on Education Policy, Washington, DC, 2008); www.cep-dc.org/index. cfm?fuseaction=document_ext.showDocumentByID&nodeID=1&DocumentID=238.
15. S. Stulich *et al.*, *National Assessment of Title I: Final Report* (U.S. Department of Education, Washington, DC, 2007); http://ies.ed.gov/ncee/pubs/20084012/.
16. Materials and methods are available as supporting material on *Science* Online.
17. California Department of Education, release 07-98, 15 August 2007; www.cde.ca.gov/nr/ne/yr07/yr07rel98.asp.
18. This research was supported by a Math Science Partnership grant from the National Science Foundation (NSF 0226948). The views presented in this paper do not necessarily reflect those of the National Science Foundation.

PHYSICS

For Quantum Information, Two Wrongs Can Make a Right

A channel too noisy to send quantum information can send secret messages, and, when combined with a similarly noisy channel, can reliably send quantum states.

Jonathan Oppenheim

Can you reliably send information down a telegraph wire that doesn't always transmit signals correctly? Claude Shannon put classical information theory on a firm footing when he showed that you can correct for transmission errors as long as there is some tiny correlation between what gets sent and what is received. What's more, Shannon quantified how much information could be reliably communicated. From its onset, classical information theory was intimately entwined with communication. The birth of quantum information theory began from an apparently different direction—cryptography—when it was realized that if you can reliably send someone quantum states, then you can use those states to exchange private messages that cannot be cracked by even the most powerful computer (*1*). This cannot be done classically without exchanging a physical key beforehand that is as long as the message you want to send. However, we are still wrestling with the corresponding question that was so central to classical information theory: How much quantum information can we reliably send down a noisy channel? On page 1812 of this issue, Smith and Yard (*2*) have discovered that we may be further from answering this question than we think, but that intriguing clues might come from the very place that initially sparked our interest in quantum information: cryptography.

Classically, a telegraph wire that is so noisy that no information can be reliably sent through it is useless. These are called zero-capacity channels. But what about the quantum case, such as trying to send information (which might be conveyed by the polarization of a single photon) through a fiber-optic cable that is so noisy it cannot be used to send any quantum state reliably?

Because our intuition tends to be classical, it was generally believed that a channel that cannot convey quantum information would also be useless. Yet a few years ago, the Horodecki brothers and I found that although these channels cannot be used to send quan-

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK. E-mail: jono@damtp.cam.ac.uk
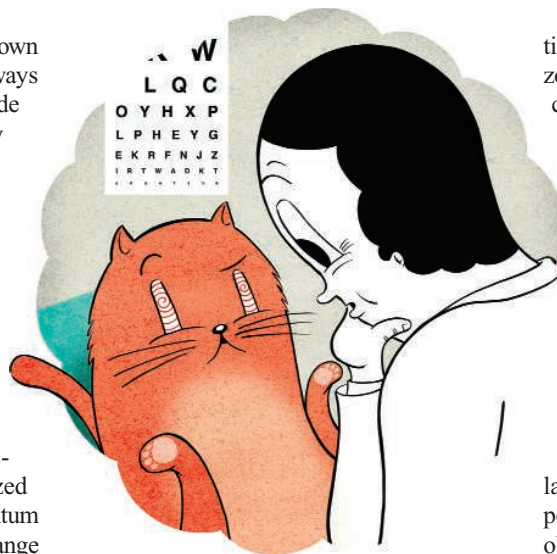
Quantum blindsight. *"You appear to be blind in your left eye and blind in your right eye. Why you can see with both eyes is beyond me..."*

tum states, they can be used to send classical private messages. Indeed, one can classify all states that, if shared over some channel, are private (*3*). What's more, this privacy is verifiable, which means that practical cryptography can be performed over these zero-capacity fibers (*4*). The belief that quantum cryptography required being able to reliably send quantum states turned out to be wrong.

Now, Smith and Yard, using results from (*5*), have shown a remarkable property of these zero-capacity quantum channels that can send private messages: They can be combined with another channel that also has zero capacity and can be used to convey quantum information. To find that two zero-capacity channels have finite capacity is a bit like finding out that $0 + 0 = 1$ (see the illustration). Each channel individually is useless for sending quantum information, but when used together, they can be used to reliably send a quantum system in any state.

Despite how perplexing this result appears from a classical perspective, there is a fairly simple way to illustrate it. Let us start with the main idea behind cryptography. Consider two parties, Alice (A) and Bob (B), who can talk on the telephone and exchange quantum states—for example, polarized photons or qubits. These are represented by vectors in a linear superposi-

tion of two states, $|0\rangle$ and $|1\rangle$, so that the horizontal polarization of a photon is $|0\rangle$, the vertical polarization is $|1\rangle$, and linear superposition can give rotations to any angle. Imagine that Alice and Bob can succeed in sharing a maximally entangled quantum state whose wave function ψ_0 can be represented as $|\psi_0\rangle = (|00\rangle_{AB} + |11\rangle_{AB})/\sqrt{2}$. In this case, Alice and Bob have their qubits in a superposition of the $|00\rangle$ quantum state and $|11\rangle$ state, with Alice (A) possessing one of the qubits and Bob (B) in possession of the other.

This $|\psi_0\rangle$ state is pure, meaning that nothing in the external world can be correlated with it. As a result, Alice and Bob can perform measurements in this $|0\rangle$, $|1\rangle$ basis and obtain a string of correlated and secret bits (their measurement outcomes will be that they each obtain 0 or each obtain 1). This string can then be used to share a private message (*6*). Any channel that can be used to share $|\psi_0\rangle$ can be used to share any other state of their choosing and is said to have positive channel capacity. Likewise, if they can share the state $|\psi_1\rangle = (|00\rangle_{AB} - |11\rangle_{AB})/\sqrt{2}$, which is also maximally entangled but has negative phase, then they can also share a private message and send quantum states.

Now, consider a channel that half of the time results in $|\psi_0\rangle$ being shared and the other half of the time results in $|\psi_1\rangle$ being shared. One can show that this channel can only send classical messages—it cannot create entanglement unless $|\psi_0\rangle$ is shared more often than $|\psi_1\rangle$ (or vice versa). To make it more interesting, the channel also sends a flag—an additional state that labels which of the two maximally entangled states has been sent. If Alice and Bob can distinguish the two flags by performing measurements on them, then they will know which entangled state they share, and they can then send private messages or quantum states as before. They can even perform a correction to the state to convert $|\psi_1\rangle$ into $|\psi_0\rangle$.

As it turns out, there exist flags that can be completely distinguished when Bob holds the entire flag, yet are arbitrarily difficult to distinguish when Alice and Bob hold different parts of the flag and must perform measurements on them in separated labs (*7*) (even if they could communicate classically with a

telephone). In such cases, they will hardly ever know whether they share $|\psi_0\rangle$ or $|\psi_1\rangle$, and their ability to send quantum states to each other is arbitrarily close to zero (one can make it exactly zero by adding small errors). However, this state is still useful for sending private messages, because Alice and Bob can still just measure as they did before in the $|0\rangle$, $|1\rangle$ basis to obtain a secret key. An eavesdropper may know whether they share $|\psi_0\rangle$ or $|\psi_1\rangle$ but not whether they obtained $|00\rangle$ or $|11\rangle$ after measurement. Thus, channels that produce these flagged states can be used to share private messages, but they cannot be used to send quantum information—they have zero quantum capacity.

Now consider another zero-capacity channel, an erasure channel, that, with probability ½, lets the quantum state through perfectly, and the rest of the time it erases the state; the receiver Bob knows an error occurred because he will measure the error state $|e\rangle$. Such a channel turns out to be useless by itself for sending quantum information, but if Alice first uses the previous zero-capacity private channel and

then puts her half of the flag down the erasure channel, then half of the time Bob can combine Alice's part of the flag that he receives from this channel with the other half that he received from the zero-capacity private channel. He can then distinguish the flag. So, half of the time, he will know whether they share $|\psi_0\rangle$ or $|\psi_1\rangle$, and he can perform a correction so that they both share the $|\psi_0\rangle$ state. This means that ¾ of the time, Alice and Bob share $|\psi_0\rangle$ instead of $|\psi_1\rangle$. This is significantly greater than half the time, and enough to create entanglement and get a positive channel capacity.

By using both the zero-capacity private channel and the zero-capacity erasure channel together, Alice and Bob can send any quantum state reliably. In the case above, the inputs that Alice sends through the two channels are not even entangled, but only classically correlated. What's more, this procedure can be easily generalized. All cryptographic protocols must distill the private states of (3), and the above protocol can be adapted to work for all of them.

This result raises many questions, not the least of which is what this work may say about

the yet unknown formula for quantum capacity. We do not know the optimal procedure for activating a private channel, whether every channel that has zero capacity (but is not classical) can have positive capacity when combined with another zero-capacity channel, or even whether every such zero-capacity channel is also a private channel. Whatever the answers, it is clear that the structure of quantum information theory is much richer than most of us ever anticipated.

References
1. C. Bennett, G. Brassard, in *Proceedings of the IEEE Conference on Computers, Systems, and Signal Processing* (IEEE, New York, 1984), pp. 175–179.
2. G. Smith, J. Yard, *Science* **321**, 1812 (2008); published online 21 August 2008 (10.1126/science.1162242).
3. K. Horodecki, M. Horodecki, P. Horodecki, J. Oppenheim, *Phys. Rev. Lett.* **94**, 160502 (2005).
4. K. Horodecki *et al.*, *Phys. Rev. Lett.* **100**, 110502 (2008).
5. G. Smith, J. A. Smolin, A. Winter, *IEEE Trans. Info. Theory* **54**, 4208 (2008).
6. A. K. Ekert, *Phys. Rev. Lett.* **67**, 661 (1991).
7. T. Eggeling, R. Werner, *Phys. Rev. Lett.* **89**, 097905 (2002).

PLANETARY SCIENCE

The Past Martian Dynamo

Numerical dynamo modeling studies may explain the observation that strong magnetic fields are only found in Mars' southern hemisphere.

Benoit Langlais[1] and Hagay Amit[2]

Measurements by Mars Global Surveyor (MGS) have revealed intense magnetic anomalies mostly located south of the crustal dichotomy, the topographic boundary separating the southern cratered highlands and the northern smooth lowlands. Assuming the dynamo of Mars was similar to that of Earth—dipolar, axial, and centered, the magnetic dichotomy implies that the magnetization of the northern hemisphere was erased at some time, and thus that the dynamo stopped operating very early in its history (1). On page 1822 of this issue, Stanley *et al.* propose an alternative model in which the dynamo is driven by a hemisphere-scale heat flux pattern at the core-mantle boundary (CMB) (2). The proposed thermal constraint is compatible with martian mantle convection models (3) and can also explain the crustal dichotomy (4). In this new scenario, the much weaker crustal magnetization in the northern hemisphere is

A magnetic dichotomy. Predicted magnetic field intensity (nT) at 300 km altitude (from (17), iso-contours are 25 nT), on top of a shaded relief of the martian surface. Northern hemisphere magnetic field anomalies are of the same order of magnitude as terrestrial magnetic field anomalies at similar altitude, and approximately one-tenth of what is measured in the southern hemisphere. The large impact craters, as well as the large volcanic provinces, show no appreciable magnetic fields at high altitude. Blue line represents the crustal dichotomy. VB, Vastitas Borealis; El, Elysium; Ol, Olympus; Is, Isidis; Th, Tharsis; Ap, Apollinaris Patera; TM, Terra Meridiani; Sy, Syrtis Major; He, Hellas; Ar, Argyre.
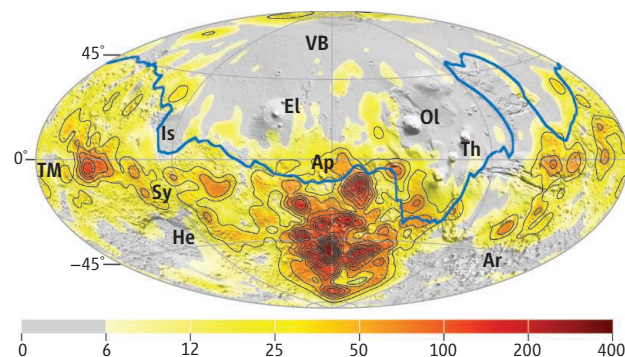
not a result of a post-dynamo process such as a giant impact (5), but rather, it was never magnetized in the first place.

Thermal core-mantle coupling can explain some features related to Earth's dynamo. Evidence suggests that the heterogeneous lower mantle affects convection and dynamo action in Earth's outer core. Paleomagnetic field models time-averaged over the past 5 million years show deviations from axial sym-

metry (6). Core flow models time-averaged over the past 150 years show persistent non-axisymmetric features (7), and the seismic properties of the upper part of Earth's inner core also exhibit an east-west hemispheric dichotomy (8).

Dynamo simulations with heterogeneous heat flux boundary conditions have been used to study the possible impact of the mantle on Earth's dynamo (9). The models successfully

[1]Laboratoire de Planétologie et Géodynamique, CNRS UMR 6112, Université de Nantes, 44322 Nantes cedex 3, France. E-mail: benoit.langlais@univ-nantes.fr [2]Équipe de Géomagnétisme, Institut de Physique du Globe de Paris, CNRS UMR 7154, 75252 Paris cedex 5, France. E-mail: hagay@ipgp.jussieu.fr

explain some of the observed non-axisymmetric features, such as the locations of the high-latitude intense geomagnetic flux patches in the modern era (10). A recent study recovered large parts of the time-averaged patterns of the paleomagnetic field, historical core flow, and inner-core buoyancy flux hemispheric dichotomy (11). In these models, core convection is primarily driven from below, whereas the variable boundary heat flux controls its long-term pattern. An upper bound for dynamo action was reported for this type of moderate heat flux anomaly amplitude (9).

Stanley et al. assume much stronger heat flux heterogeneities at the martian CMB, with convection driven from above maintaining the dynamo. Other aspects that differ from most geodynamo models (9–11) include stress-free boundary conditions and hyperdiffusivities. Two additional modeling issues concern the state of the martian core and the heat flux pattern. First, it is thought that the martian core was completely liquid during the first 500 million years (the Noachian era) (12). The effect of absence of an inner core therefore has to be evaluated. Second, the proposed dynamo model concentrates its field lines where the heat flux is the largest, i.e., below a cold downwelling mantle, whereas others (4) suggest that the thickened crust of the southern hemisphere is related to upwelling. The hemisphere-scale convection pattern in the mantle and its relationship with surface features clearly need to be better understood. Recovery of the single-hemisphere dynamo using different dynamo modeling methods and assumptions may strengthen the robustness of the proposed scenario.

Very little is known about the weak magnetic field signature of the northern lowlands. At 400 km altitude, where MGS spent most of its time, it is indeed very low, with a maximum of 20 nT above Vastitas Borealis. The absence of magnetization in the northern hemisphere (see the figure) may well be due to the single-hemisphere dynamo proposed by Stanley et al., but one can invoke other hypotheses. For example, serpentinization of the southern hemisphere lithosphere, associated with magnetite crystallization and crustal material density decrease (13), could also explain the magnetic and topographic patterns on Mars, as could rapidly varying magnetization directions resulting in null to weak fields at spacecraft altitudes.

The proposed model of Stanley et al. resolves a number of apparent discrepancies on Mars. The existence of such thermal wind dynamos may open a new avenue for dynamo modeling, for Earth but also for other planets, such as Mercury (14). As with Uranus and Neptune, and as opposed to Earth, Jupiter, and Saturn (15), the current model shows that the

past martian magnetic field was possibly non-dipolar and non-axisymmetric. Additional computations and observations are required to validate or dismiss their model. The next breakthrough will come from new observations, first when low-altitude measurements of the magnetic field are made (16) and when surface geophysical (seismic, magnetic, and heat flow) measurements are taken as planned by the European Space Agency's forthcoming Exomars rover and associated lander mission. These measurements will give some hints on the current lithosphere thickness, its origin, its relationship with possible hemisphere-scale convection, and the existence of a solid inner core. Combined with thermal evolution models, it will be possible to estimate the thermodynamic conditions on Mars during its early days. These inferences will introduce new geodynamic constraints on models of the past martian dynamo and may shed light on the reasons for its demise. The martian magnetic history is not yet over.

References and Notes
1. M. H. Acuña et al., Science **284**, 790 (1999).
2. S. Stanley et al., Science **321**, 1822 (2008).
3. H. Harder, U. R. Christensen, Nature **380**, 507 (1996).
4. J. H. Roberts, S. Zhong, Icarus **190**, 24 (2007).
5. W. S. Kiefer, Nature **453**, 1191 (2008).
6. C. L. Johnson et al., Geochem. Geophys. Geosyst. **9**, doi:10.1029/2007GC001696 (2008).
7. H. Amit, P. Olson, Phys. Earth Planet. Inter. **155**, 120 (2006).
8. F. Niu, L. Wen, Nature **410**, 1081 (2001).
9. P. Olson, U. R. Christensen, Geophys. J. Int. **151**, 809 (2008).
10. D. Gubbin et al., Phys. Earth Planet. Inter. **162**, 256 (2007).
11. J. Aubert et al., Nature **454**, 758 (2008).
12. A. J. Stewart et al., Science **316**, 1323 (2007).
13. Y. Quesnel et al., Eur. Planet. Sci. Conf. **2**, 0297 (abstr.) (2007).
14. P. J. Takley et al., Eur. Planet. Sci. Conf. **3**, 0106 (abstr.) (2008).
15. S. Stanley, J. Bloxham, Nature, **151**, 428 (2004).
16. F. Leblanc et al., Astrobiology, 10.1089/AST.2007.022 (2008).
17. B. Langlais et al., J. Geophys. Res. **109**, 10.1029/2003JE002048 (2004).
18. H.A. is supported by a grant from the Intra-European Marie Curie action.

10.1126/science.1162874

CANCER

The Metastasis Cascade

Christoph A. Klein

The view of evolution of tumor cells toward metastasis takes a new twist.

The 20th-century philosopher of science, Thomas Kuhn, proposed that when sufficient observational data accumulate that conflict with "received wisdom," the prevailing model gives way to a new paradigm (1). With the findings of Podsypanina et al. on page 1841 in this issue (2), and those of three papers published earlier this year (3–5), the field of cancer metastasis seems to be undergoing such a paradigm shift.

For decades, the metastatic dissemination of cancer has been considered the final stage in a deteriorating process. Genetic and (more recently) epigenetic changes have been thought to accumulate in the primary tumor over years before cancer cells are sufficiently mature to spread, having become "fully malignant" (see the figure). Now, Podsypanina et al. show that phenotypically normal mouse mammary epithelial cells injected into a recipient animal's bloodstream can survive at ectopic sites such as the lung, until expression

Division of Oncogenomics, Department of Pathology, University of Regensburg, Franz-Josef-Strauss-Allee 11, 93053 Regensburg, Germany. E-mail: christoph.klein@klinik.uni-regensburg.de

of the oncogenes (altered versions of normal genes) they harbor is activated, driving cell proliferation and colonization of a new site. Their findings complement work by Hüsemann et al. (4) who show that oncogene activation in mouse mammary cells triggers a genetic program, possibly governed by the transcription factor Twist, which enables dissemination of such premalignant cells from mammary tissue to lungs and bone marrow before the appearance of mammary tumors.

How do these results change our understanding of cancer metastasis? The late metastasis model faces the problem that events predisposing a cancer to metastasis are initially unselected, in that they do not provide a growth advantage over neighboring cells (6). Thus, the changes promoting metastatic dissemination are mostly associated with tumors large enough to make the change—such as a mutation—a likely event. Proponents of this model hold that by the time of migration, tumor cells harbor compound aberrations that enable their survival by inactivating programmed cell death, which would otherwise be induced by an ectopic environment. Yet Podsypanina et al. observed that apparently

normal mammary epithelial cells injected into the mouse bloodstream survived for prolonged periods in the lungs, proliferated there, and could form mammary glands when reimplanted into mammary fat tissue. Equally surprising is the number of mammary cells that survived in the lungs—as many as 1.2 per 10,000 intravenously injected cells. This is in contrast to the few metastases that arise from millions of injected tumor cells from established cell lines (7), and the observation that billions of cancer cells shunted from the peritoneal cavity of patients into their venous systems do not result in more metastases (8). Podsypanina et al. found that most of the

single disseminated tumor cells in bone marrow (10) and has been applied to both individuals suffering from various epithelial cancers and those without cancer (controls). Whereas 20 to 60% of cancer patients without manifest metastasis harbor disseminated epithelial tumor cells in bone marrow, only 2% of more than 800 controls had disseminated tumor cells, which is similar to the background noise of the assay (11). This suggests that even if resistance to programmed cell death is an inherent trait of (some) healthy epithelial cells, it is insufficient to account for cell migration. An alternative explanation, such as lack of access to

premalignant lesions (4). Twist activity also confers the ability of a cell to self-renew (5), a prerequisite for metastatic growth.

What are the consequences of this new view of the metastatic cascade for research and clinical practice? The central aspect here is not "early" versus "late" tumor cell dissemination, but the evolution of malignant cells inside versus outside the primary lesion. Collectively, the recent findings suggest a note of caution about conclusions and hypotheses based on the late metastasis model. These include the use of advanced-stage cancer cell lines for metastasis research, which might reveal nonphysiological mecha-
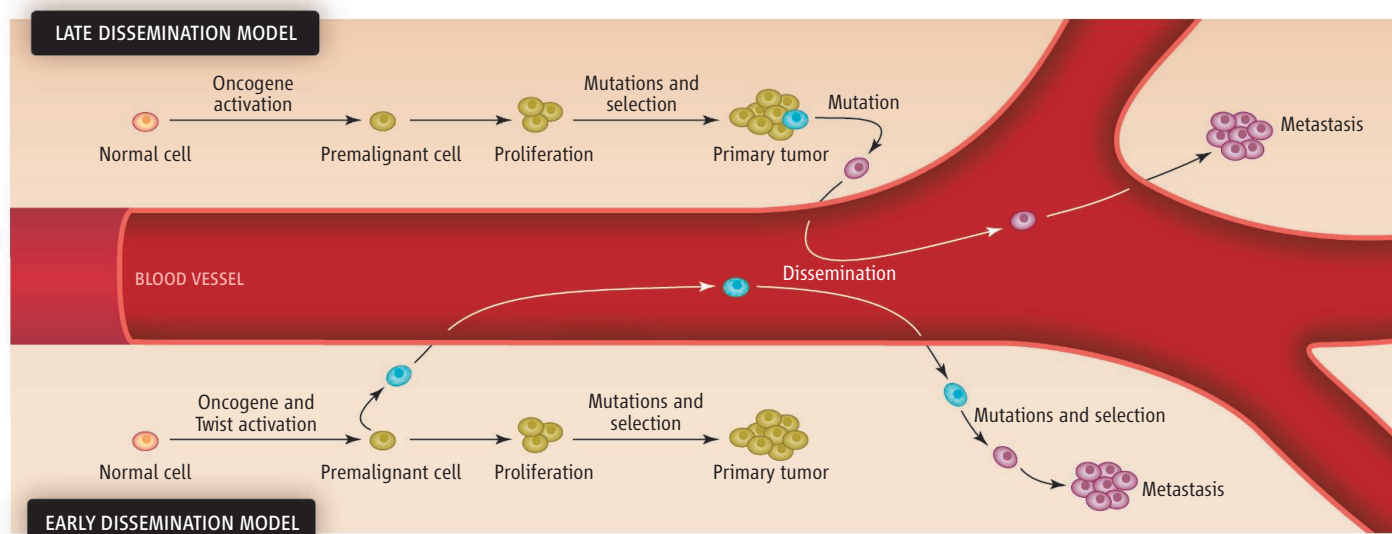
Evolution of malignancy. (Top) The late metastasis model places selection of genetic and epigenetic alterations mostly inside the primary tumor. If so, late-disseminating cells are genetically similar to the primary tumor, which can be used as a surrogate marker to choose a drug against disseminated tumor cells. **(Bottom)** By contrast, early-disseminated tumor cells accumulate such alterations at distant sites and diverge genetically from the primary tumors. Consequently, they may respond differently to drugs that are administered systemically.

surviving cells started to proliferate once the oncogenes ($Kras^{D12}$ and MYC, or $PyMT$) they carried were activated in animals.

In the context of the high frequency of some nonmalignant disorders, these results are less surprising. Endometriosis, a condition affecting 6 to 10% of women, originates from epithelial cells of the uterus that have been displaced. These untransformed epithelial cells have been found in the pelvic cavity, pericardium, pleura, aorta, and brain (9), arguably as "metastases."

The discrepancy between metastatic efficiencies of cancer cells versus healthy cells raises questions of whether advanced-stage cancer cells might be more sensitive to programmed cell death, and why dissemination of normal epithelial cells from healthy tissues to ectopic sites is not more frequent or continual. A sensitive immunoassay detects

blood vessels, must contribute a more substantial barrier to the spread. If so, actively invasive behavior is likely a critical determinant of metastasis.

If dissemination is not selected as a mutational event, it must be the consequence of a change that generates fitter cells. Activation of oncogenes increases cell proliferation, but as an initiating event, it is counteracted by senescence, a genetic program inducing irreversible growth arrest. One pathway that might be selected after oncogene activation is regulated by Twist genes, recently shown by Ansieau et al. to override senescence in murine and human cells (3). In addition, Twist regulates two genetic programs important for metastasis. It activates a cell migration-invasion program that is involved in embryonic development (called epithelial mesenchymal transition) and also in dissemination of cells from

nisms, and the assumption that therapeutic targets that are genetically activated in primary tumors will be activated in early disseminated cancer cells, small colonies of cancer cells, and metastases. The early dissemination model forces us to identify the epigenetic and genetic alterations that induce malignant development and metastatic spread. We need to distinguish changes that do initiate cell transformation from those that are merely able to. Given that cancer spread is early and selection for growth occurs at sites distant from the primary tumor, we need to establish the critical determinants of this process. If they differ from those at the primary site and lead to selection of different genotypes and phenotypes, then primary tumors present insufficient molecular markers to select "catch-all" targeting therapies. A shift of paradigm toward a model of ectopic

malignant evolution will hopefully accelerate the diagnosis, prevention, and treatment of metastatic disease.

References and Notes

1. T. S. Kuhn, in *The Structure of Scientific Theories*, F. Suppe, Ed. (Univ. of Illinois Press, Urbana, 1974), pp. 459–482.
2. K. Podsypanina *et al.*, *Science* **321**, 1841 (2008).
3. S. Ansieau *et al.*, *Cancer Cell* **14**, 79 (2008).
4. Y. Hüsemann *et al.*, *Cancer Cell* **13**, 58 (2008).
5. S. A. Mani *et al.*, *Cell* **133**, 704 (2008).
6. R. Bernards, R. A. Weinberg, *Nature* **418**, 823 (2002).
7. A. F. Chambers, A. C. Groom, I. C. MacDonald, *Nat. Rev. Cancer* **2**, 563 (2002).
8. D. Tarin *et al.*, *Cancer Res.* **44**, 3584 (1984).
9. L. C. Giudice, L. C. Kao, *Lancet* **364**, 1789 (2004).
10. G. Schlimok *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8672 (1987).
11. C. A. Klein, *Adv. Cancer Res.* **89**, 35 (2003).
12. This work was supported by a grant from the Bavarian State Ministry of Sciences, Research and the Arts.

SOCIAL SCIENCE

Unlocking the Potential of the Spoken Word

Douglas W. Oard

Advances in speech processing may soon place speech and writing on a more equal footing, with broad implications for many aspects of society.

The best available evidence suggests that the human brain, and the human facility for language, were already well developed at least by 50,000 years ago (*1*). For most of the time since then, the spoken word provided the only practical way of using language to share our understanding of the world with others. To this day, people find spoken expression and its visual correlates (such as gesture and facial expression) to be a fluid and compelling way of communicating. It was the invention of writing, however, that ignited the continuing cycle of innovation that we associate with modern society. We now stand at the threshold of a new era, one in which the spoken word can again rise to prominence.

About 5000 years ago, we see the first indications of the emergence of written language (*2*). Writing has important features that the spoken word lacks, including a degree of permanence that can help to overcome some limitations of human memory. It rapidly proliferated well beyond mere commercial records to play a multifaceted role in complex forms of social organization. This proliferation inspired other innovations: ways of finding documents again, and ways of writing that conveyed the needed context to a reader. The written word also has other attractive qualities (for example, you can read at your own pace), but permanence, findability, and contextualization are responsible for its foundational role in human civilization.

For the past century and a half, inventors have chipped away at those advantages. The earliest known recording of a human voice was made in 1860 by Edward Lyon Scott's phonautograph, although it was not until Edison's 1877 better-known phonograph that the human voice could also be reproduced using technology from the same era (*3*). Later technologies, from wire recorders through reel-to-reel tape recording, were widely adopted for commercial purposes. It was, however, introduction of the compact cassette in 1962 that ultimately made sound recording technology robust and affordable. By the end of that decade, ordinary people could record hundreds of hours of speech, for media costs of about a dollar an hour. Today, digitized speech is easily acquired (for example, using any of the world's 2.5 billion mobile phones), easily transferred over digital networks, and easily stored, all for just a few cents per hour. It would take just $100 or so of networked disk storage to record everything that you will speak or hear this year.

Digital storage is a great equalizer with regard to permanence: The same infrastructure that can reliably store digital text can equally well store digital speech. Why, then, do we not record our lives in this way? Actually, some people do. For example, researchers at Carnegie Mellon University crafted a memory aid by recording their side of conversations and then using face recognition to cue up audio from an earlier meeting—no more forgetting people's names (*4*)! Gordon Bell at Microsoft has gone further, assembling digital materials from his entire life (*5*). This works well for some things (such as e-mail), but speech is not one of them—searching through large collections of spontaneously produced speech has remained a challenge.

This situation is about to change. Commercial "media management" systems can now reliably find specific content in the well-articulated speech of news announcers, and laboratory systems can handle much of the substantial variation in speaking styles that have made automatic transcription of interviews, meetings, and telephone conversations difficult. Hardware costs are higher for speech than for born-digital text (around a factor of 100 for storage, and perhaps a factor of 1000 for processing), but it is possible today to acquire, store, and process digitized speech at lower cost than was possible for born-digital text at the dawn of the Web. Robust accommodation to noisy environments and unfamiliar words remain important challenges, however, limiting the tasks to which present speech technology can be applied.

As increasingly capable systems emerge from the laboratory, we will soon find ourselves in a world in which speech need no longer be ephemeral. How will that change

our society? No one can know for sure, but it is not difficult to envision some questions that might arise. The Carnegie Mellon system recorded only one side of the conversation because it is illegal in Pennsylvania (and 11 other U.S. states) to record full conversations without the explicit consent of all parties. Will a new balance between social costs and benefits lead us to think in more nuanced ways about when recording conversations should be permissible, just as many of us have learned to think differently about e-mail privacy at home and at the office? The wide diffusion of writing required standardization to facilitate mutual intelligibility. Will increasingly broad

College of Information Studies and Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA. E-mail: oard@umd.edu

dissemination of spoken language accelerate the demise of regional dialects and less widely spoken languages? Written contracts today have greater legal standing than verbal ones. Will that distinction persist in a world in which spoken and written words have equal permanence? How can we harness this new technology to accelerate access to new knowledge, and what would be the implications of the resulting compression of innovation cycles?

Our parents complained that our generation relied on calculators rather than learning arithmetic. Will we complain when our grandchildren rely on speech-enabled systems rather than learning to read and write? Near-universal literacy has been one of humankind's greatest accomplishments, with 82% of the world's adult population now able to read and write.

But it was the ephemeral nature of speech that gave rise to the imperative for literacy, and it is intriguing to imagine what will happen as that imperative abates. In Plato's *Phaedrus*, the Pharaoh Thamus says of writing, "If men learn this, it will implant forgetfulness in their souls: They will cease to exercise memory because they rely on that which is written" (*6*). Plato could not anticipate all the ways in which writing would be used for so much more than merely to augment memory—from an Internet that transports ideas through time and space, to great works of literature that transport our imagination to places that do not exist. What would a modern-day Plato have to say about the rise of speech to stand alongside writing as a cornerstone for our society? Our generation will unlock the full potential of the spoken word, but it may fall to our children, and to their children, to learn how best to use that gift.

References
1. P. Lieberman, *Curr. Anthropol.* **48**, 39 (2007).
2. J. Hooker, Ed., *Reading the Past: Ancient Writing from Cuneiform to the Alphabet* (British Museum Press, London, 1996).
3. J. Rosen, "Researchers play tune recorded before Edison", *New York Times*, 27 March 2008; see also www.firstsounds.org/.
4. W.-H. Lin, A. G. Hauptmann, *A Wearable Digital Library of Personal Conversations*, Joint Conference on Digital Libraries (JCDL), 2002; available at http://lastlaugh.inf.cs.cmu.edu/alex/JCDL02RememberingConversationsFinal-r1.pdf.
5. G. Bell, J. Gemmell, *Sci. Am.* **296**, 58 (March 2007).
6. Plato, Phaedrus 275a, quoted from *Platos's Phaedrus*, trans. with introduction and commentary by R. Hackford (Cambridge Univ. Press, Cambridge, UK., 1952).

ECONOMICS

Can Neural Data Improve Economics?

Researchers are exploring how neurobiology can guide economic experiments and refine economic models.

Eric Maskin

Modern neuroimaging techniques—functional magnetic resonance imaging (fMRI), positron emission tomography scans, and so on—allow us to peer inside the brain and see what is going on when experimental subjects make economic decisions such as how to bid in auctions. The data on, say, dopamine release in the nucleus accumbens, or—as Delgado *et al.* (*1*) report on page 1849 of this issue—blood oxygen in the striatum, are certainly fascinating in their own right. But can they improve our understanding of economic behavior?

Opinions diverge on this question. Neuroeconomists Camerer *et al.* recently predicted that "We will eventually be able to replace the simple mathematical ideas that have been used in economics with more neurally-detailed descriptions" (*2*). By contrast, economic theorists Gul and Pesendorfer maintain that neuroscience evidence is irrelevant to economics because "the latter makes no assumptions and draws no conclusions about the physiology of the brain" (*3*). Limited to current practice in economics, the Gul-Pesendorfer assertion is correct. In a standard economic model, a decision-maker is confronted with several options, and the purpose of the exercise is to predict which one the subject will select. The model assumes and asserts nothing about the subject's brain states, nor is there any call for it to do so as long as the prediction is accurate. But predictions based on standard choice models are sometimes far from satisfactory, and so in principle, we might improve matters by allowing predicted behavior in the model to depend not only on the economic options but also on neurophysiological information.

So far, the field of neuroeconomics has not developed such an expanded model. Moreover, even when it does so, there are knotty problems of obtrusiveness and privacy to be resolved before one could perform brain scans outside the laboratory. The field has been moving quickly enough so that there is cause for optimism that all this will ultimately transpire, but integrating neural information into everyday economics is probably a good many years off.

What can be done with brain scans before that happy time? One possibility advocated by Delgado *et al.* is to use them for discriminating among standard economic models, in which neurophysiological variables (such as changes in blood oxygen levels) do not

Buying behavior. Why do people overbid for items at auction?

appear. Most puzzling economic phenomena admit quite a few conceivable alternative explanations, and neural data can streamline the process of finding the best one—suggesting follow-up experiments or new hypotheses. The authors use this approach to try to illuminate subjects' behavior in high-bid auction experiments. While they are probably right about how neural data can be useful, their application of this principle to auctions does not seem entirely successful.

In a high-bid auction, each potential buyer for the item being sold makes a sealed bid (i.e., quotes an amount of money without disclosing that amount to the other buyers). The buyer making the highest bid wins the item

School of Social Science, Institute for Advanced Study, Einstein Drive, Princeton, NJ 08540, USA. E-mail: maskin@ias.edu

and pays the seller that bid. High-bid auctions call for strategic behavior by buyers. If the item is worth v to a buyer, she will bid strictly less than v, because bidding her actual valuation would gain her nothing: She would get something worth v but also pay v. How much she "shades" her bid—that is, bidding below what the item is worth to her—will depend on what she expects others will do. Game theory predicts that each buyer will bid so as to maximize her expected payoff, given that all other buyers do the same. The result is what is called an equilibrium.

In one of the Delgado *et al.* experiments, there are two buyers, whose assigned valuations for the item being sold are drawn independently from a uniform distribution on the numbers between 0 and 100. If the buyers are risk-neutral—that is, if a buyer's expected payoff is her net gain from winning (valuation minus bid) times the probability of winning—then in equilibrium, the buyer will bid half her valuation. However, Delgado *et al.* found—as have many other similar experiments—that subjects generally bid more than this: They "overbid."

Delgado *et al.* discuss two standard explanations for overbidding. One is that subjects are risk-averse rather than risk-neutral—they strictly prefer the expectation of a monetary gamble to the gamble itself. The other is that they get an extra psychic benefit from beating out another buyer. What the authors do not mention, however, is that both hypotheses are now considered somewhat dubious: Recent experimental evidence seems in conflict with each of them (*4*). Thus, it is welcome that Delgado *et al.* propose their own explanation, based on fMRI studies they performed.

Unfortunately, it is not completely clear what this new hypothesis is. The fMRI data show that subjects experience a lower blood oxygen level in the striatum in response to losing an auction, but no significant change in reaction to winning one. The authors interpret this result as suggesting that subjects experience "fear of losing" and that this fear accounts for their overbidding. But actually modeling fear explicitly—making it precise—does not seem straightforward.

A natural modeling device would be simply to subtract something from the subject's payoff when she loses. However, such a modification would not accord with the authors' findings in their subsequent experiment. In the follow-up, there were two treatments: one in which a subject is initially given a bonus sum of money S but told that she has to return it if she loses the auction; the other in which the subject is promised that if she wins she

will get S. The two treatments are, *ex post*, identical: In both cases, the subject ends up with the bonus if and only if she wins. However, in practice, subjects bid more in the former treatment than the latter. Such behavior sharply contradicts the "payment subtraction" hypothesis, under which behavior in the two treatments would be the same. Moreover, it seems difficult to find a natural alternative formulation of the "fear of losing" idea that explains the results simultaneously from both Delgado *et al.* experiments. Even so, there is a well-known principle that could account for the behavioral discrepancy between the two treatments in the follow-up experiment: the "endowment" effect (*5*). When a subject is given a bonus S at the outset, she may become possessive and so move more aggressively to retain it than she would act to obtain a contingent bonus at the end of the experiment.

As for why subjects overbid, perhaps the answer is that high-bid auctions are just too complex for a typical buyer to analyze completely systematically. The buyer will easily see that she has to shade her bid (bid strictly below v) to get a positive payoff. Still, she won't want to shade too much because shading reduces her probability of winning. A

simple rule of thumb would be to shade just a little. But this leads immediately to overbidding, because risk-neutral equilibrium bidding entails a great deal of shading: A buyer will bid only one-half her valuation.

In short, Delgado *et al.*'s discovery of a dip in striatal blood oxygen levels when buyers lose in an auction is an intriguing neurophysiological finding, although it is not so clear that it has yet led to a better economic model of buyers' behavior. Still, the philosophy of Delgado *et al.*—that neural findings show great potential for improving economic analysis—is one that should be endorsed, well before the time when neuroscience and economics become one.

References and Notes
1. M. R. Delgado, A. Schotter, E. Y. Ozbay, E. A. Phelps, *Science* **321**, 1849 (2008).
2. C. Camerer, G. Loewenstein, D. Prelec, *J. Econ. Lit.* **43**, 9 (2005).
3. F. Gul, W. Pesendorfer, "The case for mindless economics," www.princeton.edu/~pesendor/mindless.pdf (2005).
4. J. Kagel, D. Levin, "Auctions: A survey of experimental research, 1995–2008," www.econ.ohio-state.edu/kagel/Auctions_Handbook_vol2.pdf (2008).
5. R. Thaler, *J. Econ. Behav. Org.* **1**, 39 (1980).
6. I thank NSF for research support.

CHEMISTRY

Nonlinear Thinking About Molecular Energy Transfer

Richard M. Stratt

Although solvent molecules move about randomly in a liquid, an experiment showed that changing their initial arrangement affected the rate of a chemical process.

In the movies, nobody cares what the extras are doing or saying, but you would notice if they were missing. Chemical reactions in solution are similar. The solvent molecules need to be there to ferry energy into and out of the reacting molecules, but when chemists study how molecules change into one another in chemical reactions, solvent molecules barely show up in the credits. In fact, the working hypothesis of most studies of chemical reactions run in solution is that the details of how the reaction funnels energy into the solvent tend to average out: The ability to transfer energy depends on the solution's friction (its intrinsic ability to absorb energy), not on precisely how the energy is donated. This

Department of Chemistry, Brown University, Providence, RI 02912, USA. E-mail: Richard_Stratt@brown.edu

notion, which sanctions not having to remeasure or recalculate results with every tiny shift in reaction conditions, receives its justification from linear response theory, an idea that is used in many fields to understand complex systems. Thus, the observation that linear response does not always work as expected, as Bragg *et al.* (*1*) demonstrate on page 1817 of this issue for the simplest chemical reaction—shifting an electron—is striking.

A basic tenet of linear response theory is that the energy flow in macroscopic systems is proportional to whatever causes it, with the proportionality constant a measure of the relevant friction. Linear response theory accounts for current being proportional to voltage in Ohm's law, for example (with the resistance a constant, reflecting the constancy of the friction) (*2*). However, in more recent

applications to ultrafast laser studies of chemical events in solution, these same ideas are being applied, with remarkable success, at a molecular level (*3*).

This molecular connection occurs because chemical reactions are energy converters. A typical reaction rearranges the internal energy stored in chemical bonds, along with the external energy available from molecular translational kinetic energy, into the energy needed to create other bonding patterns.

solute, such as its charge). These studies showed that the linear response theory could account for transfers of large quantities of energy, even though one might have thought that linearity would rely on getting small effects out of small perturbations (*2*). More modern perspectives tend to attribute linearity to the consistency of the random Gaussian fluctuations that a solute usually sees when surrounded by a macroscopic number of other molecules (*10*, *11*). Although nonlinear

energy relaxation (*13*, *16*) (yet another example of nonlinearity), that such a void is likely to be a much more significant perturbation to the solvent's fluctuations than a minor local excess solvent density (such as that induced by adding an electron). The failure of linear response theory here may not be all that different from that expected theoretically.

What these special instances of nonlinear energy transfer do promise, however, are some long-sought experimental entries into the microscopic origins of friction. The slowing down of mechanical motion by friction is nothing but the effectively irreversible transfer of molecular kinetic energy into a sea of countless other molecular degrees of freedom. The second law of thermodynamics guarantees that such transfers occur, but it cannot explain how or when they happen, or reveal the specific molecular motions involved. Linear response theory in this case is the villain, as it hides all of this detail in quantities set by the

Do the details matter? Whether an electron is added to Na$^+$ (**Left**) or removed from Na$^-$ (**Right**), the immediate outcome is a neutral Na atom surrounded by solvent molecules (open circles) in unfavorable locations. Although the resulting solvent arrangements differ appreciably, both situations relax to a more favorable arrangement (**Middle**) within tens of picoseconds. Linear response theory predicts that this relaxation is determined solely by the solvent fluctuations shown in the middle panel, meaning that the two solvation pathways in red should exhibit identical dynamics, hiding the differing details of their starting points. Bragg *et al.* showed that the right-hand side pathway is noticeably slower.

However, there is often energy left over, and this residual energy relaxes into molecular vibrational and translational energy.

In even the best studies of this energy relaxation in chemical reactions, experiments keep track of relatively few of the actors (*4*). The precise molecular energies that come into play after a chemical reaction in solution are not easy to mimic in the laboratory, but some recent experiments have deposited as much as 40 kJ/mol of energy into a single, well-defined quantum transition of a C-H or an O-H stretching vibration and obtained spectroscopic signatures of energy flowing from place to place within (*5*) and between (*6*) molecules during the subsequent relaxation. Nonetheless, these energies are not close to the 200 to 300 kJ/mol that is often rearranged in chemical reactions. Worse, different ways of inserting the same amounts of energy into a molecule could, in principle, lead to different dynamics (*7*).

Given the many uncertainties, it would be convenient to have linear response theory to rely on to say that many of these details do not matter. Indeed, some of the first computer simulations of energy relaxation confirmed the accuracy of linear response predictions at the molecular level in solvation processes (*3*, *8*, *9*), the essence of which is the energy lowering induced by rearranging solvent molecules around a newly created solute (or, more frequently, around a newly altered state of a

responses were seen in a few simulations of solvation (*12*), they could all be attributed to specific variations in these Gaussian fluctuations triggered by gross changes in the surrounding solvent structure (*13*).

What Bragg *et al.* developed was an elegant experiment to look for nonlinearities in solvation. The usual experimental protocol has been to electronically excite a dissolved dye molecule and interpret the evolution of its fluorescence with time (*3*). Instead, Bragg *et al.* generated a ground-state neutral sodium atom (Na) in tetrahydrofuran (THF) by two different electron-transfer routes—ejecting an electron from a sodium anion (Na$^-$) and shifting an electron from a nearby iodide ion to a sodium cation (Na$^+$) (see the figure). Had the subsequent rearrangements of THF molecules exhibited a linear response, the electronic absorption spectra of the neutral Na products would have had identical patterns of time evolution, despite their different starting points. The observation of appreciable differences seems to show that nonlinearity can appear even with the solvation changes encountered when a single electron is transferred (*14*).

Still, removing an electron from a single, reasonably small, atomic anion does create a hole in the liquid that is fairly substantial on the scale of the anion. We know, both from previous work from Schwartz's group (*15*) and from more recent experimental and theoretical studies of high-energy rotational

materials involved (the Ohm's law resistance, for example) and not by the specifics of the motion. These new experiments offer welcome instances when the obscuring fog of microscopic linear response lifts, revealing those missing molecular details.

References and Notes
1. A. E. Bragg, M. C. Cavanagh, B. J. Schwartz, *Science* **321**, 1817 (2008).
2. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford Univ. Press, New York, 1987), chap. 8.
3. R. M. Stratt, M. Maroncelli, *J. Phys. Chem.* **100**, 12981 (1996).
4. C. G. Elles, F. F. Crim, *Annu. Rev. Phys. Chem.* **57**, 273 (2006).
5. Z. Wang, A. Pakoulev, D. Dlott, *Science* **296**, 2201 (2002).
6. J. C. Deàk, Y. Pang, T. D. Sechler, Z. Wang, D. Dlott, *Science* **306**, 473 (2004).
7. S. Yan, Y.-T. Wu, B. Zhang, X.-F. Yue, K. Liu, *Science* **316**, 1723 (2007).
8. M. Maroncelli, *J. Chem. Phys.* **94**, 2084 (1991).
9. E. A. Carter, J. T. Hynes, *J. Chem. Phys.* **94**, 5961 (1991).
10. B. B. Laird, W. H. Thompson, *J. Chem. Phys.* **126**, 211104 (2007).
11. P. L. Geissler, D. Chandler, *J. Chem. Phys.* **113**, 9759 (2000).
12. T. Fonseca, B. M. Ladanyi, *J. Mol. Liq.* **60**, 1 (1994).
13. G. Tao, R. M. Stratt, *J. Chem. Phys.* **125**, 114501 (2006).
14. J. S. Bader, D. Chandler, *Chem. Phys. Lett.* **157**, 501 (1989).
15. C. J. Smallwood, W. B. Bosma, R. E. Larsen, B. J. Schwartz, *J. Chem. Phys.* **119**, 11263 (2003).
16. A. C. Moskun, A. E. Jailaubekov, S. E. Bradforth, G. Tao, R. M. Stratt, *Science* **311**, 1907 (2006).
17. The author's work is supported by the NSF (CHE-0809385).

EDUCATION

Innovations Liven Up Undergraduate Science Classes

A comic book may seem like an odd choice for assigned reading in a college class, but biology professor Jay Hosler has transformed the medium into a resource for his undergraduate students. Where comics are best known for telling the stories of superheroes and goofy teenagers, *Clan Apis* and other Hosler comics instead explore the evolution of the eye, the stages of cell division, and the life cycle of the honey bee.

"Comic books are an example of a great innovative education model to reach students who are nervous or not confident in the science classroom," said Hosler, explaining his second life as comic author and illustrator at Juniata College in Huntington, Pennsylvania.

Hosler's color-splashed pages were among the intriguing resources discussed at the 2008 Course, Curriculum, and Laboratory Improvement (CCLI) conference held 13 to 15 August in Washington, D.C.

Clan Apis follows the life cycle of a bee named Nyuki.

The event, cosponsored by AAAS, brought together more than 500 stakeholders in an effort to transform science, technology, engineering, and mathematics (STEM) teaching on college campuses across the country.

Since 1999, the National Science Foundation (NSF) has distributed grants to colleges and universities through the CCLI program, encouraging them to adopt a more innovative approach to STEM teaching. Last year, NSF provided around $67.5 million supporting 262 new initiatives on 203 campuses around the country.

AAAS's Education and Human Resources (EHR) division has been a partner in the initiative since organizing the first CCLI conference in 2004. This year's event is in keeping with AAAS's mission to improve STEM curricula and help science faculty collaborate with their colleagues across disciplines, said EHR Deputy Director Yolanda George.

"Undergraduate courses are usually the last chance we get to help citizens and future leaders understand the nature of science and the importance of STEM to our lives—as individuals and as members of the larger society," said EHR Director Shirley Malcom. "We can't emphasize enough how important it is to improve education at this level."

The meeting's participants shared their own innovative lesson plans, which included a wall-climbing robot with gecko feet, the materials engineering of a snow ski, and geometry concepts illustrated by the students' own drawings and sculptures.

Another way that the CCLI program has encouraged science teaching for nonscience majors is to "focus more on public and civic issues such as sustainability and the ethics of genetic engineering and the use of nanotechnology in consumer products," said Myles Boylan, one of the directors of the CCLI program. "Science has so much to say about this, and students find it very motivating."

Many of Hosler's students are interested in becoming primary or elementary school teachers, but they "are scared out of their minds about science," he said. "And if they are afraid of science, there is no way they could teach it well. We need to show them that they can not only learn about it but, eventually, teach it to the next generation."

Jeanne Narum, director of Project Kaleidoscope, an informal alliance to improve undergraduate STEM education, said colleges and universities are sometimes hesitant to try new educational approaches because the ability of the new models to increase student comprehension is untested or not widely understood.

Evaluating new STEM methods is a critical part of the CCLI program, according to Boylan. "We've heard it loud and clear from our grantees: 'Help us do a better job of assessing our projects,'" he said.

Robert Grossman, a professor of chemistry at the University of Kentucky, said faculty members might be able to convince their university's administrators to implement innovative programs by showing that they are in the institution's best interest.
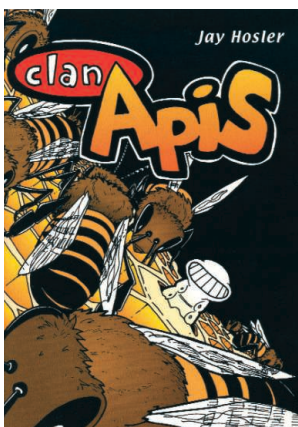
Several years ago, Grossman said, his institution discovered that students receiving low or failing grades in chemistry were more likely to withdraw from the course or college as a whole, at great expense to the university.

In an effort to increase student success, a colleague of Grossman used a CCLI grant to develop ChemExcel, a peer-led tutoring program for undergraduates in chemistry. Grossman said the University of Kentucky eagerly supported the program after it was shown to reduce student failure rates.

—Benjamin Somers

INTERNATIONAL

An Emerging Engagement on Rwandan Education

A high-level AAAS delegation pledged to work with top Rwandan science and education leaders on development of science-related curricula during 2 days of meetings in the central African nation.

The talks on 27 and 28 August focused on Rwanda's plan to use science and technology to drive economic growth and build human capital as it continues to recover from the 1994 genocide that left nearly 1 million people dead.

Throughout the meetings, Rwandan Education Minister Daphrosa Gahakwa, Science Minister Romain Murenzi, and other key S&T officials stressed that education will be critical to the plan's success. Representing AAAS were Shirley Malcom, head of Education and Human Resources; Vaughan Turekian, chief international officer; and Sarah Banas, program associate in the International Office.

"During our visit we were again reminded of the central role that Rwanda's leadership is placing on science and technology in their development strategy," said Turekian. "This provides a very good opportunity for AAAS and other science-based organizations to share experiences and work together to help build capacity."

Malcom said AAAS will provide Rwandan officials with science education resources developed by Project 2061, the association's pioneering science literacy initiative.

Over the past year, AAAS and Rwandan leaders have built a promising relationship. Last October, AAAS Board Chair David Baltimore visited Rwandan President Paul Kagame and other officials. Kagame traveled to Boston in February to deliver a plenary address at the AAAS Annual Meeting.

CREDIT: JAY HOSLER

AAAS Annual Election: Preliminary Announcement

The 2008 AAAS election of general and section officers will be held in November. All members will receive a ballot for election of the president-elect, members of the Board of Directors, and members of the Committee on Nominations. Members registered in more than one section will receive ballots for elections for each section they are enrolled in.

Candidates for all offices are listed below. Additional names may be placed in nomination for any office by petition submitted to the Chief Executive Officer no later than October 26. Petitions nominating candidates for president-elect, members of the Board, or members of the Committee on Nominations must bear the signatures of at least 100 members of the Association. Petitions nominating candidates for any section office must bear the signatures of at least 50 members of the section. A petition to place an additional name in nomination for any office must be accompanied by the nominee's curriculum vitae and statement of acceptance of nomination.

Biographical information for the following candidates will be enclosed with the ballots mailed to members in October.

Slate of Candidates

GENERAL ELECTION

President: Alice Huang, California Institute of Technology; Harold Mooney, Stanford Univ.

Board of Directors: Jerry Melillo, Marine Biological Laboratory; Julia Phillips, Sandia National Laboratories; David Sabatini, NYU Medical Ctr.; Maria Elena Zavala, California State Univ., Northridge.

Committee on Nominations: Steven Chu, Lawrence Berkeley National Laboratory; Jack Dixon, Howard Hughes Medical Institute; Jonathan Dordick, Rensselaer Polytechnic Institute; Steven Fienberg, Carnegie Mellon Univ.; Paul Friedman, UCSD Medical Center; M.R.C. Greenwood, UC-Davis; Susan Hackwood, California Council on Science & Technology; Sallie Keller-McNulty, Rice Univ.

SECTION ELECTIONS

Agriculture, Food, and Renewable Resources

Chair Elect: Harry Klee, Univ. of Florida; Brian Larkins, Univ. of Arizona

Member-at-Large of the Section Committee: Barbara Valent, Kansas State Univ.; Jeff Volenec, Purdue Univ.

Electorate Nominating Committee: Charles Brummer, Univ. of Georgia; Candace Haigler, North Carolina State Univ.; Ann Hirsch, UCLA; Mark Sorrells, Cornell Univ.

Council Delegate: Daniel Cosgrove, Pennsylvania State Univ.; Stanley Roux, Univ. of Texas, Austin

Anthropology

Chair Elect: Clark Larsen, Ohio State Univ.; Christopher Ruff, Johns Hopkins Univ.

Member-at-Large of the Section Committee: Carol Ward, Univ. of Missouri; Sarah Williams-Blangero, Southwest Foundation for Biomedical Research

Electorate Nominating Committee: Daniel Brown, Univ. of Hawaii; Dolores Piperno, Smithsonian Institute; Dawnie Wolfe Steadman, Binghamton Univ.; Anne C. Stone, Arizona State Univ.

Astronomy

Chair Elect: Alan P. Boss, Carnegie Institution of Washington; Mario Livio, Space Telescope Science Institute

Member-at-Large of the Section Committee: Jack O. Burns, Univ. of Colorado; Donald Campbell, National Astronomy and Ionosphere Ctr.

Electorate Nominating Committee: Giovanni G. Fazio, Harvard Smithsonian Ctr. for Astrophysics; Arlo U. Landolt, Louisiana Univ.; James E. Neff, College of Charleston; Michael Werner, Jet Propulsion Laboratory

Atmospheric and Hydrospheric Sciences

Chair Elect: Alan Robock, Rutgers Univ.; Donald Wuebbles, Univ. of Illinois

Member-at-Large of the Section Committee: Leo Donner, Geophysical Fluid Dynamics Laboratory; Kevin E. Trenberth, National Center for Atmospheric Research

Electorate Nominating Committee: Jim Coakley, Oregon State Univ.; Qiang Fu, Univ. of Washington; Margaret (Peggy) Lemone, National Center for Atmospheric Research; Mark H. Thiemens, UCSD

Biological Sciences

Chair Elect: Trudy Mackay, North Carolina State Univ.; James R. Spotila, Univ. of Arkansas

Member-at-Large of the Section Committee: Nipam H. Patel, UC-Berkeley; Margaret Werner-Washburne, Univ. of New Mexico

Electorate Nominating Committee: Chris T. Amemiya, BRI, Virginia Mason Medical Ctr.; Joan W. Bennett, Rutgers Univ.; Judith Berman, Univ. of Minnesota; Eric Ursell Selker, Univ. of Oregon

Chemistry

Chair Elect: Charles Casey, Univ. of Wisconsin-Madison; Dale Poulter, Univ. of Utah

Member-at-Large of the Section Committee: Peter Wipf, Univ. of Pittsburgh; Ronald Woodard, Univ. of Michigan College of Pharmacy

Electorate Nominating Committee: Jonathan Ellman, UC-Berkeley; Charles Craik, UC-San Francisco; Brian Stoltz, California Institute of Technology; Amos B. Smith, III, Univ. of Pennsylvania

Dentistry and Oral Health Sciences

Chair Elect: Francis Macrina, Virginia Commonwealth Univ.; Margarita Zeichner-David, Univ. of Southern California

Member-at-Large of the Section Committee: Richard Lamont, Univ. of Florida; Ira B. Lamster, Columbia Univ. College of Dental Medicine

Electorate Nominating Committee: Peter Ma, Univ. of Michigan School of Dentistry; Laurie McCauley, Univ. of Michigan School of Dentistry; Frank Scannapieco, School of Dental Medicine Univ. at Buffalo; Robert G. Quivey Jr., Univ. of Rochester

Education

Chair Elect: Joe Krajcik, Univ. of Michigan; Mary Nakkhleh, Perdue Univ.

Member-at-Large of the Section Committee: Angelo Collins, Knowles Science Teaching Foundation; Jay B. Labov, National Academy of Science

Electorate Nominating Committee: Mary Atwater, Univ. of Georgia; Linda Froschauer, Weston Public Schools; Jon D. Miller, Michigan State Univ.; Suzanne O'Connell, Wesleyan Univ.

Engineering

Chair Elect: H. Vincent Poor, Princeton Univ.; Duncan T. Moore, The Institute of Optics

Member-at-Large of the Section Committee: Christine Maziar, Univ. of Notre Dame; Jerome Schultz, UC-Riverside

Electorate Nominating Committee: Rena Bizios, Univ. of Texas, San Antonio; Kathy Ferrara, UC-Davis; Kristen Fichthorn, Univ. of Pennsylvania; Pradeep Khosla, Carnegie Mellon Univ.

Council Delegate: Jose Cruz, Ohio State Univ.; Gail H. Marcus, Consultant; James L. Merz, Univ. of Notre Dame; C.D. (Dan) Mote Jr., Univ. of Maryland

General Interest in Science and Engineering

Chair Elect: Charles Lytle, North Carolina State Univ.; Kathryn D. Sullivan, Ohio State Univ.

Member-at-Large of the Section Committee: Sharon M. Friedman, Lehigh Univ.; Alexander Polonsky, Marine Hydrophysical Institute

Electorate Nominating Committee: Suzanne Gage Brainard, Univ. of Washington; Robert Griffin, Marquette Univ.; Marilee Long,

Colorado State Univ.; Gloria J. Takahashi, La Habra High School

Geology and Geography
Chair Elect: Malcolm Hughes, Univ. of Arizona; Stephen Jackson, Univ. of Wyoming
Member-at-Large of the Section Committee: Sally P. Horn, Univ. of Tennessee; Jean Lynch-Stieglitz, Georgia Institute of Technology
Electorate Nominating Committee: Elizabeth Canuel, College of William & Mary; Eugene Domack, Hamilton College; Timothy Fisher, Univ. of Toledo; David Stahle, Univ. of Arkansas

History and Philosophy of Science
Chair Elect: Richard Creath, Arizona State Univ.; Jeffery L. Sturchio, The Merck Company Foundation *Member-at-Large of the Section Committee*: Robert Brandon, Duke Univ.; Heather Douglas, Univ. of Tennessee
Electorate Nominating Committee: Mark Largent, James Madison College; Nancy Nersessian, Georgia Institute of Technology; Peter Railton, Univ. of Michigan; Alain Touwaide, Smithsonian Institution
Council Delegate: Jane Maienschein, Arizona State Univ.; Virginia Trimble, UC-Irvine

Industrial Science and Technology
Chair Elect: Jennie C. Hunter-Cevera, Univ. of Maryland Biotechnology Institute; Vijayan Nair, Univ. of Michigan
Member-at-Large of the Section Committee: Manuel Gomez, Univ. of Puerto Rico; Harry S. Hertz, NIST
Electorate Nominating Committee: Robert Boily, Inforex Inc.; Quinghuang Lin, IBM; John Pizzonia, Fujifilm Life Science's Applications Laboratory; Harold Schonhorn, retired
Council Delegate: Orlando Auciello, Argonne National Laboratory; Steven Popper, The Rand Corporation

Information, Computing, and Communication
Chair Elect: Bart Selman, Cornell Univ.; Manuela Veloso, Carnegie Mellon Univ.
Member-at-Large of the Section Committee: Julia Gelfand, UC-Irvine; John (Jack) Hill, USGS
Electorate Nominating Committee: Christine Borgman, UCLA; Bonnie Carol, Information International Associates; Casimir Kulikowski, Rutgers Univ.; William Woods, ITA Software

Linguistics and Language Science
Chair Elect: David W. Lightfoot, Georgetown Univ.; Thomas Wasow, Stanford Univ.
Member-at-Large of the Section Committee: Suzanne Flynn, MIT; Mark Liberman, Univ. of Pennsylvania

Electorate Nominating Committee: William J. Poser, Yinka Dene Language Institute; Edward Stabler, UCLA; Elizabeth Traugott, Stanford Univ.; Douglas H. Whalen, Haskin Laboratories **Mathematics**
Chair Elect: Kenneth Millett, UC-Santa Barbara; Williams Velez, Univ. of Arizona
Member-at-Large of the Section Committee: Tony F. Chan, National Science Foundation; Carl C. Cowen, Indiana University-Purdue Univ.
Electorate Nominating Committee: Douglas Arnold, Univ. of Minnesota; Jonathan Borwein, Dalhousie Univ.; Wade Ellis, West Valley Community College; Robert M. Fossum, Univ. of Illinois, Urbana-Champaign

Medical Sciences
Chair Elect: Gary A. Koretsky, Univ.of Pittsburgh School of Medicine; Judy Lieberman, Harvard Medical School
Member-at-Large of the Section Committee: Robert Doms, Univ. of Pennsylvania; Rino Rappuoli, Chiron Corporation
Electorate Nominating Committee: Wendy Brown, Washington State Univ.; Beverly Davidson, Univ. of Iowa; Phil Greenberg, Univ. of Washington; Thomas B. Nutman, National Institutes of Health
Council Delegate: Etty (Tika) Benveniste, Univ. of Alabama, Birmingham; Terence Dermody, Vanderbilt Univ.; James M. Hughes, Emory Univ.; Marcelo Jacobs-Lorena, John Hopkins School of Public Health; Jennifer M. Puck, UC-San Francisco; Reed Pyeritz, Univ. of Pennsylvania; Gwendalyn J. Randolph, Mount Sinai School of Medicine; Douglas Richman, UC-San Diego; Paul Rothman, Univ. of Iowa; Thomas Wellems, National Institutes of Health

Neuroscience
Chair Elect: Nominees to be announced
Member-at-Large of the Section Committee: Richard Huganir, John Hopkins Univ.; Gail Mandel, Oregon Health & Science Univ.
Electorate Nominating Committee: Erik D. Herzog, Washington Univ.; Frank LaFerla, UC-Irvine; Michael S. Wolfe, Harvard Medical School/Brigham and Women's Hospital; Tony Wyss-Coray, VA Palo Alto Health Care System

Pharmaceutical Science
Chair Elect: Kenneth Thummel, Univ. of Washington; Gary Pollack, Univ. of North Carolina
Member-at-Large of the Section Committee: William Beck, Univ. of Illinois; David Ross, Univ. of Colorado
Electorate Nominating Committee: Per Artursson, Uppsala Univ.; Kenneth Brouwer, Qualst; Donald E. Mager, Univ. at Buffalo; Craig K. Svensson, Purdue Univ.

Physics
Chair Elect: S. James Allen, UC-Santa Barbara; Charles W. Clark, NIST
Member-at-Large of the Section Committee: Richard F. Casten, Yale Univ.; Alexander L. Fetter, Stanford Univ.
Electorate Nominating Committee: Patrick D. Gallagher, NIST; Elizabeth H. Simmons, Michigan State Univ.; Michael Witherell, UC-Santa Barbara; Ali Yazdani, Princeton Univ.

Psychology
Chair Elect: Judy Deloache, Univ. of Virginia; Stephen Suomi, National Institutes of Health
Member-at-Large of the Section Committee: Randolph Blake, Vanderbilt Univ.; Jenny Saffran, Univ. of Wisconsin
Electorate Nominating Committee: Randall Engle, Georgia Tech; Denise Park, Univ. of Texas, Dallas; Barbara Rolls, Pennsylvania State Univ.; David Shapiro, UCLA
Council Delegate: Bennett I. Bertenthal, Indiana Univ.; John Gabrieli, MIT

Social, Economic, and Political Sciences
Chair Elect: Eugene Rosa, Washington State University; Cora B. Marrett, National Science Foundation
Member-at-Large of the Section Committee: Ronald J. Angel, Univ. of Texas, Austin; Wendy Baldwin, Population Council
Electorate Nominating Committee: Anil Deolalikar, UC-Riverside; Don C. Des Jarlais, Albert Einstein College of Medicine; Howard Leventhal, Rutgers Univ.; Robert F. Rich, Univ. of Illinois
Council Delegate: Nicholas Christakis, Harvard Medical School; Guillermina Jasso, New York University

Societal Impacts of Science and Engineering
Chair Elect: Bruce Lewenstein, Cornell Univ.; Stephen D. Nelson, Virginia Tech. Univ.
Member-at-Large of the Section Committee: Kevin Finneran, Issues in Science and Technology; Melanie Leitner, Prize4Life
Electorate Nominating Committee: Lida Anestidou, The National Academies; Ezra Heitowit, Universities Research Association, Inc.; Robert M. Simon, US Senate Committee on Energy & Natural Resources; Michael L. Telson, Univ. of California System

Statistics
Chair Elect: Joel B. Greenhouse, Univ. of Pittsburgh; Jessica Utts, UC-Irvine
Member-at-Large of the Section Committee: Charmaine Dean, BC Cancer Agency; Kenneth W. Wachter, UC-Berkeley
Electorate Nominating Committee: Paul P. Biemer, Univ. of North Carolina; May A. Foulkes, George Washington Univ.; Subir Ghosh, UC-Riverside; Nancy Reid, Univ. of Toronto

Assembling Materials with DNA as the Guide

Faisal A. Aldaye, *et al.*
Science **321**, 1795 (2008);
DOI: 10.1126/science.1154533

Assembling Materials with DNA as the Guide

Faisal A. Aldaye,[1] Alison L. Palmer,[2] Hanadi F. Sleiman[1]*

DNA's remarkable molecular recognition properties and structural features make it one of the most promising templates to pattern materials with nanoscale precision. The emerging field of DNA nanotechnology strips this molecule from any preconceived biological role and exploits its simple code to generate addressable nanostructures in one, two, and three dimensions. These structures have been used to precisely position proteins, nanoparticles, transition metals, and other functional components into deliberately designed patterns. They can also act as templates for the growth of nanowires, aid in the structural determination of proteins, and provide new platforms for genomics applications. The field of DNA nanotechnology is growing in a number of directions, carrying with it the promise to substantially affect materials science and biology.

Today, we can chemically synthesize complex molecules such as palytoxin, vitamin B12, or Taxol with remarkable angstrom-scale precision and fabricate intricately designed micron-scale electronic components at the rate of billions per second. These advances may suggest that we have conquered most major frontiers of chemical construction. Yet, nature illustrates that we are nowhere near the limit of exquisite control over organization; it possesses an extraordinary capacity to assemble complex nanostructures with active and specialized functions. Our ability to precisely position components on the nanometer scale the way nature does, and to do so in a parallel rather than a serial manner, is still limited and is a key goal in nanotechnology and materials science.

Self-assembly, the spontaneous association of components into organized structures using noncovalent interactions, is the chief method that nature uses to achieve complexity. Of the natural self-assembling molecules, DNA is arguably the most remarkable. A cooperative interplay of hydrogen-bonding, π-stacking, electrostatic, and hydrophobic interactions drives one DNA strand to assemble with its complement into a double helix according to extremely precise base-pairing rules. Additional attributes, such as rigidity on the nanoscale, a diameter of ~2 nm, and a near-infinite number of potential sequences, extend DNA's reach beyond a genetic blueprint for life. DNA is emerging as an attractive tool for nanoscience as well; it is a highly promising template for organizing nanomaterials in a programmable way. Research in this area promises to allow us to use DNA to dictate the precise positioning of materials and molecules into any deliberately designed structure, thus approaching the effortless manner in which nature generates complexity and function.

Structural DNA Nanotechnology

By exploiting DNA's structural features and powerful base-pairing rules, the field of structural DNA nanotechnology aims to generate nanopatterned materials and to control motion at the nanoscale. Its initial challenge was to convert the one-dimensional (1D) DNA molecule into 2D and 3D structures. Seeman and his research group looked to nature's branched DNA structures (for example, the Holliday junction) to meet this challenge. By assembling four DNA strands into a stable four-way junction and incorporating single-stranded "sticky ends" at the periphery for hybridization, Seeman developed an artificial branched DNA "tile" (Fig. 1A, top) (1). The second major challenge of DNA nanotechnology was to generate more rigid junctions, which are essential to achieve well-defined 2D DNA assemblies. By joining two DNA double helices with a single strand that begins on one helix and switches onto an adjacent helix, Seeman generated tiles that have "crossovers" and addressable sticky ends at their edges and are of greater rigidity (Fig. 1B, left). The group used these tiles to construct well-defined 2D lattices with predesigned periodicity (2).

These principles of construction have since been used, adapted, and developed to generate systems with very fine control over design and function (2). A number of basic structural motifs have been designed that can be classified as planar tiles, branched junctions, or helix bundles. Planar tiles are formed from several parallel helices joined by crossover junctions (Fig. 1B, left) and were used to synthesize linear arrays, 2D lattices, and DNA nanotubes. Branched junctions are constructed with multiple DNA arms that radiate from a focal point and are held together with crossover junctions to minimize flexibility (Fig. 1B, middle). They have been used to generate 2D arrays with square, hexag-onal, and compound cavities. Helix bundles are tiles constructed by joining parallel DNA helices that are not coplanar (that is, they cooperatively produce a curvature) by using multiple crossover junctions (Fig. 1B, right). They have been assembled into extended 2D arrays and well-defined DNA nanotubes. These structural motifs (2) collectively provide a toolbox to rationally access a rich number of 2D DNA architectures and refine DNA materials assembly. It is of note that in structural DNA nanotechnology, DNA is used to provide all the parameters for self-assembly: connectivity, structural features, and programmability.

In this approach, DNA tiles are typically made using strands with different sequences to prevent the formation of undesired structures. In practice, however, this requires synthesizing a large number of components and mixing these in exact stoichiometric ratios for successful assembly. By incorporating identical sequences (sequence symmetry) in DNA strands, Mao found that a stable four-way junction can be constructed from three strands instead of nine and that it assembles into the desired square grid array with an increased long-range order (2). Yan used sequence symmetry to tackle the problem of constructing finite arrays, rather than extended 2D assemblies, from a small number of DNA tiles (2). Thus, judicious incorporation of sequence symmetry in DNA strands merely used as architectural elements, such as struts and junctions, can simplify tile-based assembly. Winfree proposed the concept of "algorithmic DNA self-assembly" to increase complexity in DNA assembly. This was achieved by designing a set of DNA building blocks that represent "Wang tiles." Conceptually, Wang tiles contain a single color on each of their four sides and assemble so that the adjacent sides of each square are of the same color. This requirement necessarily means that each tile can fit in a specific manner within the assembly. Winfree adapted this methodology to construct rectangular-shaped DNA tiles, with four addressable sticky ends at each side as Wang tiles, and demonstrated the feasibility of assembling these according to a set of algorithmic rules initiated by a nucleating strand. Impressively complex fractal patterns can be generated using this approach, from a minimal set of DNA tiles (3).

In addition to tiles created by simple Watson-Crick base-pairing of DNA, many other nucleic acid motifs have been developed. For example, Jaeger showed that folded RNA molecules could be assembled together like a jigsaw puzzle (Fig. 1C) (4); Willner synthesized a polycatenated DNA ladder as a mechanically interlocked system onto which proteins, nanoparticles, and dyes were fixed with precise control (Fig. 1D) (5); and Sen demonstrated the "synapsing" together of two DNA duplexes into a ladderlike structure through guanine-quadruplex formation (Fig. 1E) (6).

[1]Department of Chemistry, McGill University, 801 Sherbrooke Street West, Montreal, QC H3A 2K6, Canada. [2]Canadian Institute for Advanced Research, 180 Dundas Street West, Suite 1400, Toronto, ON M5G 1Z8, Canada.

*To whom correspondence should be sent. E-mail: hanadi.sleiman@mcgill.ca

Two-dimensional DNA templates provide the opportunity to template the positioning of materials with nanoscale precision. For example, nanoparticle assemblies are promising components for functional devices. Their collective properties, such as electron transport, optical coupling, and magnetic interactions, depend on their relative arrangement; thus, the DNA-mediated control of nanoparticle organization promises to greatly affect the fields of nanoelectronics and nanooptics, among others. Sequence-specific DNA-templated 1D organization of gold nanoparticles was demonstrated by Alivisatos by labeling gold nanoparticles with DNA strands that determined their exact position onto complementary single-stranded DNA templates (*7*). Kiehl (*8*), Seeman (*9*), and Yan (*10*) showed the sequence-encoded organization of gold nanoparticles into well-defined rows on a number of 2D DNA lattices, demonstrating remarkable control over periodicity and arrangement. In addition to nanoparticles, the organization of proteins on DNA templates can possibly lead to "enzyme factories" and substrates for proteomics and can also shed light on the nature of protein-protein interactions. The biotin-avidin interaction was used by LaBean and Yan to generate an alternating assembly of streptavidin molecules onto a 2D square DNA array (Fig. 1F) (*11*). Antigen-antibody interactions enabled Mao (*12*)

and Yan and Chaput (*13*) to organize antibodies to fluorescein and c-*myc*, respectively, on antigen-modified DNA arrays. Aptamer-protein interactions are particularly interesting because it is possible to discover aptamers for any protein by using the systematic evolution of ligands by exponential enrichment (SELEX), which means that these interactions can potentially be adapted to organize any protein. For example, Yan used a 2D DNA array, modified with two different aptamers, to assemble two proteins into alternating lines with no unwanted cross-talk (Fig. 1G) (*14*). Although the previous examples require labeling the DNA array with molecules that recognize the materials to be patterned (for example, DNA strands, biotin, antigens, and aptamers), unmodified DNA arrays can also be used to template materials assembly. The Dervan lab developed a class of polyamides that sequence-selectively bind the minor groove of DNA. These molecules can selectively bind to 2D DNA arrays and, when modified with biotin, they mediate the organization of streptavidin into lines with control over sequence and periodicity (*15*).

A number of applications of materials assembled by structural DNA nanotechnology are already starting to emerge in both biotechnology and materials science. Precisely positioned 2D DNA and protein nanoarrays (rather than conventional microarrays) can be useful in many areas of genomics, proteomics, diagnostics, and tissue engineering. By assembling different DNA tile arrays, each with a specific recognition molecule and "bar-coded" with a specific dye, Yan developed a platform that allows the simultaneous detection of multiple biological analytes. This method may be simpler than DNA or protein microarrays for small-scale profiling of bioanalytes (*16*). Mao employed a 2D DNA array as a reusable "mask" to create 2D gold nanopatterns via vapor deposition into the array's cavities (*17*). This method is promising for controlling topography in the nanoscale regime at a much higher resolution than conventional photolithography. Turberfield showed the binding of the protein RuvA to the four-way junctions of a DNA 2D lattice. This resulted in a 2D crystalline array of this protein, which allowed for its structural elucidation by using cryogenic transmission electron microscopy with a resolution of 30 Å. Interestingly, this DNA-binding protein was found to dramatically modify the geometry of the DNA motif by changing the structural features of these junctions when bound to them (*18*).

DNA Origami

In "DNA origami," a single continuous strand of DNA is systematically folded using a large number of smaller DNA strands. This approach was first reported by the group of Joyce, who synthesized a single continuous DNA strand that is 1.6 kb long and, in a single step, annealed it in the presence of five smaller strands to generate a DNA octahedron (*19*). Ingeniously, Rothemund
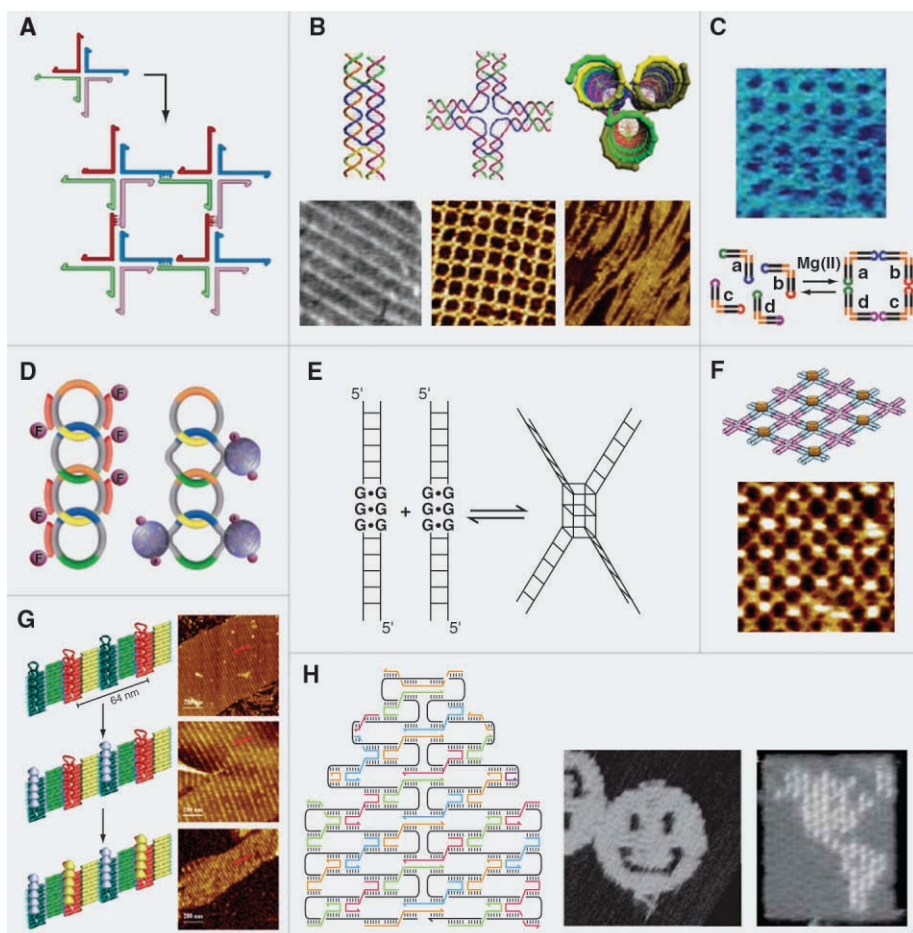
Fig. 1. (**A**) Four DNA strands assemble into a four-way junction with sticky ends, which can further assemble into 2D structures. (**B**) Motifs in structural DNA nanotechnology. Tiles (left) can assemble into periodic 2D arrays; junctions (middle), such as this four-way junction, can result in a 2D square lattice; and helix bundles (right), such as this triple bundle, can generate a 1D DNA nanotube. (**C**) Folded RNA molecules are used to selectively construct extended RNA arrays with different cavity sizes and shapes. (**D**) Interlocked DNA circles form catenated ladder-assemblies, containing single-stranded sides for molecule organization (left) or for protein binding when folded into aptamers (right). F, fluorophore. (**E**) Guanine tracks incorporated into duplexes can be used to snap four strands into a four-way junction composed of a guanine quadruplex. (**F**) A four-arm junction modified with biotin is used to organize streptavidin onto a periodic square array. (**G**) Four tiles are assembled into a 2D array with alternating rows of two different aptamers to organize two different proteins into alternating lines. (**H**) In DNA origami, a long DNA strand is folded into the desired structure and is held into shape using many stapling strands (left). This approach is used to access different 2D architectures (middle) and to draw 2D shapes, such as the map of the Western Hemisphere (right).

generalized this approach to fold naturally occurring genomic DNA into any 2D shape (*20*). In his DNA origami approach, the long strand is folded into the desired shape by a number of smaller strands ("stapling strands") (Fig. 1H, left). The sequences of these strands are computationally designed. Rothemund was able to assemble the same initial long strand of DNA into squares, rectangles, stars, smiley faces (Fig. 1H, middle), and many other 2D shapes. The power of this approach lies in its addressability: Because each stapling strand is a unique sequence, each strand is also a spatially addressable bit. Hairpins were incorporated into stapling strands to write words, such as "DNA," and to draw complex objects, such as the outline of the Western Hemisphere (Fig. 1H, right). DNA origami will be useful for accessing larger DNA shapes with highly addressable surfaces. In a recent application, Yan constructed origami-based DNA nanoarrays for label-free RNA detection of the three genes *Rag*-1, c-*myc*, and β-*actin* (*21*). Shih, Chou, and Douglas synthesized DNA nanotubes by "rolling" a DNA origami array and used the alignment made by these liquid crystalline materials to measure nuclear magnetic resonance parameters in transmembrane proteins (*22*), a powerful way to use DNA organization in protein structure determination.

Supramolecular DNA Assembly

Supramolecular chemistry exploits intermolecular forces to control the organization of organic and inorganic assemblies. After 40 years of research, it has generated a toolbox of molecular components that assemble with a high degree of control (*23*). These components are structurally rigid, geometrically diverse, and intrinsically functional (for example, are photoactive or redox active or possess magnetic properties), in contrast to the more passive DNA branch points. Now a new research area is evolving that brings the tools of supramolecular chemistry and DNA nanotechnology together. "Supramolecular DNA assembly" blends rationally designed DNA building blocks with synthetic organic and inorganic molecules, which give structural and functional advantages both to the initial self-assembly process and to the final construct.

One exciting potential of incorporating synthetic molecules into DNA is that they can dramatically influence the structure of assemblies and introduce different motifs in DNA nanotechnology. Bergstrom (*24*), Shchepinov (*25*), and von Kiedrowski (*26*) presented branched DNA structures with organic corner units that self-assemble into well-defined nanostructures. Because these structures contained identical DNA strands, however, mixtures of DNA assemblies were obtained. Sleiman developed DNA building blocks, containing two arms of different sequences, and a rigid organic corner unit, which selectively assembled into a discrete DNA hexagon (*27*). This approach was used to organize six gold nanoparticles into a hexagon (Fig. 2A),

thus providing model systems to study single-electron transport and optical coupling in gold nanoparticle assemblies.

Synthetic molecules can bring a number of additional interactions into DNA nanotechnology. For example, replacing the DNA bases with supramolecular building blocks can expand the genetic alphabet: self-complementary iso-guanines, with two hydrogen-bonding faces that are oriented at an angle that forms a pentameric assembly, result in a higher-order DNA pentaplex rather than the classical duplex (*28*). Incorporating extended aromatic molecules as connectors of DNA strands allows folding of these strands through π-π stacking (DNA "foldamers") (Fig. 2B) (*29*), and replacing the termini of DNA strands with ligands allows metal coordination to override base-pairing and loop DNA into a cycle (*30*). Attaching a polymer to the end of DNA can cause microphase separation, resulting in DNA micellar aggregates (Fig. 2C) (*31*). Synthetic molecules can also covalently link DNA structures. They have been used to create, for example, parallel DNA helix bundles with porphyrins at their cores (*32*) and to "snap" together DNA-modified organic conjugated modules into tailor-made conjugated assemblies (*33*).

Another important impact of incorporating synthetic molecules (for example, transition metals) is that they can give much needed function to the passive DNA scaffolds. Metal complexes can be photoactive and electroactive and can possess magnetic and catalytic properties. In contrast to growing materials on the exterior of a DNA

strand, incorporating transition-metal complexes into DNA can create pure and monodisperse DNA structures with preserved self-assembly capabilities. Two approaches have been investigated. The first designs nucleobases for complexing transition metals. Shionoya incorporated five consecutive copper-DNA base pairs into a DNA duplex to create a copper stack likened to a self-assembling DNA nanomagnet (Fig. 2D) (*34*) and later, with Carell, created DNA multimetallic stacks with two selectively incorporated transition metals (*35*).

The second approach, which uses metal complexes as vertices, has allowed for coordination geometries, bond angles, and functionalities unavailable to carbon compounds. Sleiman reported the synthesis of metal-linked branched-DNA building blocks, and assembled a cyclic metal-DNA nanostructure with luminescent metal vertices and DNA arms (*36*). Han constructed a DNA triangle with three iron-corner units and three DNA arms (*37*), and McLaughlin showed the synthesis of a ruthenium complex with six DNA arms (*38*). However, because these approaches require using a small subset of completely unreactive metal complexes, very few additional metals have been incorporated as vertices. Sleiman recently incorporated a range of reactive transition metals into DNA junctions, using a method that allows the DNA duplexes and transition metal complexes to synergistically stabilize each other (Fig. 2E) (*39*). Many applications can be expected for metal DNA nanostructures in such areas as multimetallic catalysis, sensing, artifi-
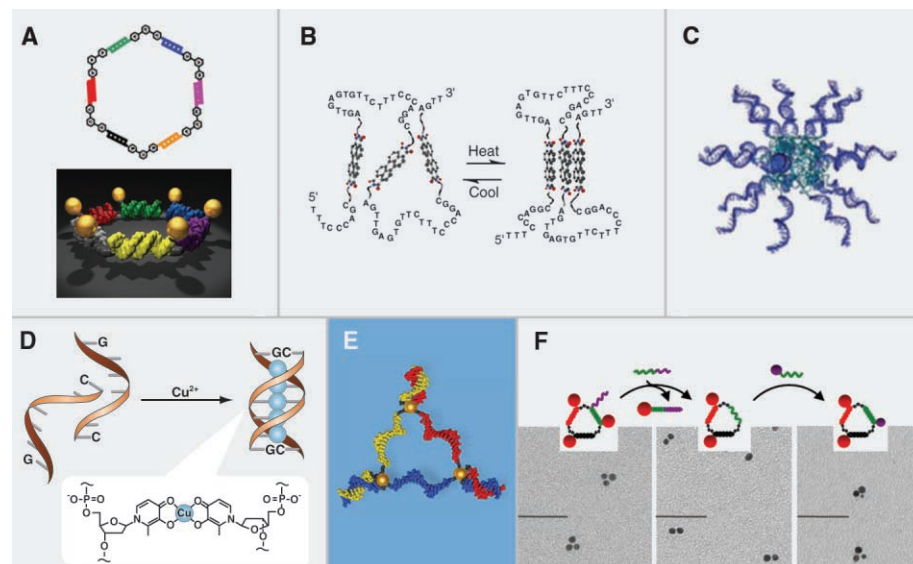
Fig. 2. (**A**) A cyclic DNA hexagon is selectively assembled from building blocks composed of two DNA arms and an organic junction and is used to organize six gold nanoparticles into a 2D discrete hexagon. (**B**) Incorporating extended aromatic molecules into DNA creates folded structures through π-stacking. (**C**) DNA strands attached to polymers result in block copolymers that assemble into micelles via microphase separation of the incompatible blocks. (**D**) Replacing DNA bases with coordinating hydroxypyridone ligands generates copper stacks within DNA strands. (**E**) A metal-DNA nanostructure: a DNA triangle with iron bis(terpyridine) vertices. (**F**) A write/erase experiment using discrete gold nanoparticle assemblies from single-stranded and cyclic DNA templates with organic vertices. A single particle is selectively removed using a fully complementary strand and is replaced with a differently sized particle.

cial photosynthesis, data storage, nanooptics, and nanoelectronics.

In supramolecular DNA assembly, synthetic molecules can contribute to the connectivity and the structure of the final molecule, and DNA is used as the programmable component. One direct consequence of this is that the structures no longer need to be double-stranded and interwoven with crossover units but can now be single-stranded and dynamic. Sleiman reported the synthesis of single-stranded and cyclic DNA structures with rigid organic corners and used them as dynamic scaffolds to organize gold nanoparticles (40) with the ability to write/erase and structurally switch these assemblies upon addition of added specific DNA strands (Fig. 2F). The group then constructed 3D DNA cages capable of switching and changing their size between three predefined states (41). There have been many elegant designs for DNA nanomachines that respond to specifically added DNA strands or other molecules (42). All these examples have led to the development of molecule-responsive DNA materials. In contrast to other environmentally responsive materials (for example, materials that change with pH, light, or oxidation), these allow for the selective control of different parts within the same device and for the incorporation of many molecular triggers to cause individual changes, thus increasing our ability to communicate and externally manipulate structures.

Three-Dimensional Assembly

Three-dimensional structures made of DNA have tremendous potential to encapsulate and release drugs, regulate the folding and activity of encapsulated proteins, selectively encage nanomaterials, and assemble 3D networks for catalysis and biomolecule crystallization. Seeman reported early examples of 3D DNA objects with the topology of a cube (43) (Fig. 3A, middle) and a truncated octahedron. Joyce reported the synthesis of an octahedron (19) (Fig. 3A, top), and Turberfield generated a rigid and chiral DNA tetrahedron (Fig. 3A, bottom), in which the group encapsulated the protein cytochrome c (44). More recently, new methods that increase the range of 3D structures and their ease of synthesis have been reported. Sleiman developed a face-centered approach to 3D DNA construction (Fig. 3B). By breaking down 3D objects into discrete 2D DNA shapes, such as triangles, squares, pentagons, and hexagons with organic vertices, a large number of 3D DNA cages were
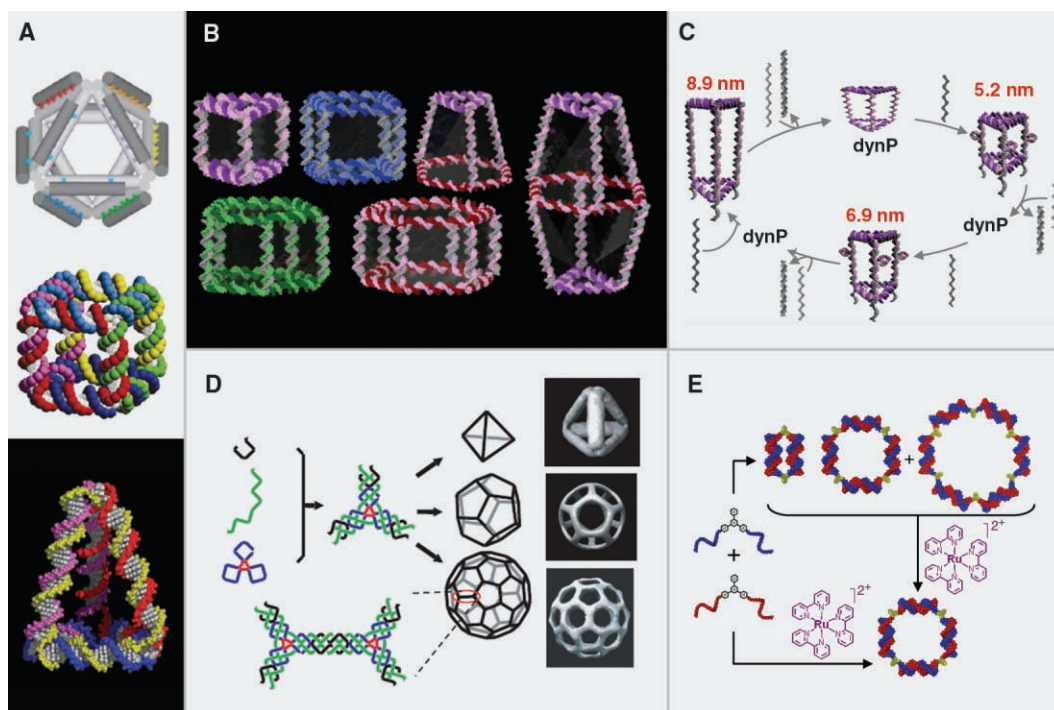
Fig. 3. (**A**) A DNA octahedron (top), cube (middle), and tetrahedron (bottom). (**B**) Single-stranded and cyclic DNA triangles, squares, pentagons, and hexagons with organic vertices as their corner units are assembled into 3D triangular prisms, cubes, pentameric and hexameric prisms, a heteroprism, and a biprism. (**C**) 3D DNA assemblies generated using this approach can also be dynamic. A triangular prism is switched between three predefined lengths using a series of strands capable of rigidifying and erasing. dynP, dynamic prism. (**D**) A symmetric three-arm junction assembles into a 3D tetrahedron, octahedron, or buckyball. Access to the desired structure is determined by the flexibility and concentration of the symmetric junction. (**E**) The molecule Ru(bpy)²⁺ affects the self-assembly outcome of two symmetric DNA building blocks and selectively mediates the assembly of a single product, a DNA square.

quantitatively accessed. These include a triangular prism, a cube, a pentameric and hexameric prism, a heteroprism, and a biprism (41). The approach also allowed for the construction of the first dynamic 3D DNA capsule, whose size was switched reversibly between three different lengths (Fig. 3C) (41). Mao adopted the rules of symmetry to access 3D assemblies from building blocks with identical arms. By controlling the flexibility within a symmetric three-arm junction, the group synthesized a tetrahedron, a dodecahedron, and a buckyball from a minimal set of building blocks (Fig. 3D) (45). Turberfield also showed dynamic switching of a DNA tetrahedron (44), and von Kiedrowski reported the synthesis of a DNA dodecahedron with organic vertices (46).

In a different approach, Mirkin (47) and Gang (48) created 3D gold nanoparticle crystals with long-range order without using a preassembled DNA template. Instead, they modified gold nanoparticles with single-stranded DNA, which allowed them to control the interparticle distances and packing dynamics, and induced crystallization.

Current Challenges

These examples illustrate the power and promise of DNA as a template to precisely position materials on the nanoscale. In order to move

this research forward, two challenges need to be addressed: the correction of errors that arise in DNA assembly, and the replication and scale-up of DNA nanostructures.

As the complexity of DNA assemblies increases, so will the number of the DNA sequences required to form them. This will necessitate using overlapping, degenerate strands that may assemble into undesirable products. Biological systems have developed a number of elegant strategies to proofread and remove errors during and after assembly. Inspired by these systems, Lu used an approach in which deoxyribozymes (DNAzymes) specifically locate and cleave misassembled structures in gold nanoparticle assemblies (49). In the presence of the "correct" DNA strands, the DNAzyme is not properly folded and is inactive; however, in the presence of the "incorrect" DNA strands, the DNAzyme is properly folded and proceeds to cleave and remove the errors. Using the rules of dynamic combinatorial chemistry, Sleiman used an external molecule to proofread and correct for errors (50). DNA building blocks with identical arms initially generated a library of many assemblies, but adding the small molecule Ru(bpy)₃²⁺, an electrostatic binder of DNA, forced the library to converge to only one member, a DNA square (Fig. 3E). Pierce generated a number of metastable folded intermediates that systematically

interact with each other in a cascading approach to generate the final product (*51*). By programming the biomolecular pathway leading to the final assembly, and not just the final product, Pierce's approach prevents error formation.

To enable practical applications, it is necessary to develop techniques that copy, amplify, and scale up the synthesis of these structures. Assemblies generated using DNA origami are constructed from a single long DNA strand and, in principle, could readily be amplified using the already existing biological machinery. Seeman and Yan recently reported the successful enzymatic amplification of a building block found in classical DNA nanotechnology: a branched DNA tile containing several double cross-over motifs (*52*).

Assemblies generated using supramolecular chemistry can also be amplified. Von Kiedrowski showed the chemical copying of a molecule composed of three single-stranded DNA arms that branch out from a single organic vertex (*53*). Ultimately, today's materials are synthesized in large amounts using chemical approaches. It is therefore worthwhile to further investigate chemical methods to economically amplify these molecules. In principle, as long as the base-pair information within DNA is preserved, different backbones could also be used. Lynn copied the information from DNA into a daughter molecule, with preserved base sequence and a synthetic oligomeric backbone (*54*). The development of economically feasible amplification methods will allow for the structures developed by DNA nanotechnology to be widely used in materials science.

Conclusions and Outlook

DNA's simple code forms our genetic blueprint for life. But the field of DNA nanotechnology has invited us to look at the code in a whole new way: as a means to precisely position materials. This code can now help dictate the specific location of materials and the structure of assemblies, creating linear, 2D and 3D assemblies. It can also control motion, creating capsules that expand and contract, molecular "walkers" that move directionally along a track, and tweezers that grab desired targets (*55*). The DNA code can even dictate specific mechanistic pathways, enabling DNA origami to fold and DNA hairpins to open each other sequentially.

Nature builds complexity in a hierarchical way. It progressively increases length scales and relies on a number of noncovalent interactions, including DNA base-pairing, to drive assembly. Supramolecular DNA assembly is a means to weave in principles of hierarchical complexity and new interactions into DNA nanostructures, and opens the door to assembling more diverse functional structures with greater ease.

The next step will be to investigate the possibilities for making practical materials with DNA nanotechnology. DNA's ability to guide patterning of transition metals, nanoparticles, and proteins into deliberate designs gives it tremendous potential for answering many important challenges in science. For instance, is it now conceivable to assemble an artificial photosynthesis system, create a multienzyme catalytic "factory," or access complex nanoelectronic circuitry, using DNA? Can DNA be employed to generate combinatorial patterns of precisely positioned small-molecule ligands that probe cooperative binding and allosteric interactions in proteins and aid in the discovery of new multivalent drugs? Can DNA cages be used as biodegradable and molecule-responsive materials for the specific on-demand delivery of drugs to diseased cells? Can 3D DNA crystalline arrays be created and used as designer molecular hosts to template protein crystallization, as initially anticipated by Seeman, or to induce catalysis and new chemistry within their cavities?

These are only a handful of the opportunities created by our remarkable control over the organization of materials using DNA. It is by identifying the important challenges of biology, chemistry, physics, engineering, and medicine that we can put these patterned nanomaterials to the test and evolve this exciting field into an applied, central area of research.

References and Notes

1. N. C. Seeman, *Nature* **421**, 427 (2003).
2. C. Lin, Y. Liu, S. Rinker, H. Yan, *Chem. Phys. Chem.* **7**, 1641 (2006) and references therein.
3. P. W. Rothemund, N. Papadakis, E. Winfree, *PLoS Biol.* **2**, 2041 (2004).
4. A. Chworos *et al.*, *Science* **306**, 2068 (2004).
5. Y. Weizmann, A. B. Braunschweig, O. I. Wilner, Z. Cheglakov, I. Willner, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5289 (2008).
6. R. P. Fahlman, D. Sen, *J. Am. Chem. Soc.* **124**, 4610 (2002).
7. A. P. Alivisatos *et al.*, *Nature* **382**, 609 (1996).
8. K. Y. Pinto *et al.*, *Nano Lett.* **5**, 2399 (2005).
9. J. Zheng *et al.*, *Nano Lett.* **6**, 1502 (2006).
10. J. Sharma, R. Chhabra, Y. Liu, Y. Ke, H. Yan, *Angew. Chem. Int. Ed.* **45**, 730 (2006).
11. H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, T. H. LaBean, *Science* **301**, 1882 (2003).
12. Y. He, Y. Tian, A. E. Ribbe, C. Mao, *J. Am. Chem. Soc.* **128**, 12664 (2006).
13. B. A. R. Williams, K. Lund, Y. Liu, H. Yan, J. C. Chaput, *Angew. Chem. Int. Ed.* **46**, 3051 (2007).
14. R. Chhabra *et al.*, *J. Am. Chem. Soc.* **129**, 10304 (2007).
15. J. D. Cohen, J. P. Sadowski, P. B. Dervan, *J. Am. Chem. Soc.* **130**, 402 (2008).
16. C. Lin, Y. Liu, H. Yan, *Nano Lett.* **7**, 507 (2007).
17. Z. Deng, C. Mao, *Angew. Chem. Int. Ed.* **43**, 4068 (2004).
18. J. Malo *et al.*, *Angew. Chem. Int. Ed.* **44**, 3057 (2005).
19. W. M. Shih, J. D. Quispe, G. F. Joyce, *Nature* **427**, 618 (2004).
20. P. W. K. Rothemund, *Nature* **440**, 297 (2006).
21. Y. Ke, S. Lindsay, Y. Chang, Y. Liu, H. Yan, *Science* **319**, 180 (2008).
22. S. M. Douglas, J. J. Chou, W. M. Shih, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6644 (2007).
23. J.-M. Lehn, *Chem. Soc. Rev.* **36**, 151 (2007).
24. J. Shi, D. E. Bergstrom, *Angew. Chem. Int. Ed. Engl.* **36**, 111 (1997).
25. M. S. Shchepinov, K. U. Mir, J. K. Elder, M. D. Frank-Kamenetskii, E. M. Southern. *Nucleic Acids Res.* **27**, 3035 (1999).
26. M. Scheffler, A. Dorenbeck, S. Jordan, M. Wüstefeld, G. von Kiedrowski, *Angew. Chem. Int. Ed.* **38**, 3311 (1999).
27. F. A. Aldaye, H. F. Sleiman, *Angew. Chem. Int. Ed.* **45**, 2204 (2006) and references therein.
28. J. C. Chaput, C. Switzer, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10614 (1999).
29. W. Wang, W. Wan, H.-H. Zhou, S. Niu, A. D. Q. Li, *J. Am. Chem. Soc.* **125**, 5248 (2003).
30. M. Goritz, R. Kramer, *J. Am. Chem. Soc.* **127**, 18016 (2005).
31. K. Ding, F. E. Alemdaroglu, M. Borsch, R. Berger, A. Herrmann, *Angew. Chem. Int. Ed.* **46**, 1172 (2007).
32. M. Endo, N. C. Seeman, T. Majima, *Angew. Chem. Int. Ed.* **44**, 6074 (2005).
33. K. V. Gothelf, A. Thomsen, M. Nielsen, E. Clo, R. S. Brown, *J. Am. Chem. Soc.* **126**, 1044 (2004).
34. K. Tanaka, A. Tengeiji, T. Kato, N. Toyama, M. Shionoya, *Science* **299**, 1212 (2003).
35. K. Tanaka *et al.*, *Nat. Nanotechnol.* **1**, 190 (2006).
36. D. Mitra, N. Di Cesare, H. F. Sleiman, *Angew. Chem. Int. Ed.* **43**, 5804 (2004).
37. J. S. Choi *et al.*, *J. Am. Chem. Soc.* **126**, 8606 (2004).
38. K. M. Stewart, J. Rojo, L. W. McLaughlin, *Angew. Chem. Int. Ed.* **43**, 5808 (2004).
39. H. Yang, H. F. Sleiman, *Angew. Chem. Int. Ed.* **47**, 2443 (2008).
40. F. A. Aldaye, H. F. Sleiman, *J. Am. Chem. Soc.* **129**, 4130 (2007).
41. F. A. Aldaye, H. F. Sleiman, *J. Am. Chem. Soc.* **129**, 13376 (2007) and references therein.
42. The development of DNA machines has been an exciting and productive area of research; for a recent review see (*55*).
43. J. H. Chen, N. C. Seeman, *Nature* **350**, 631 (1991).
44. R. P. Goodman *et al.*, *Nat. Nanotechnol.* **3**, 93 (2008) and references therein.
45. Y. He *et al.*, *Nature* **452**, 198 (2008).
46. J. Zimmermann, M. P. J. Cebulla, S. Monninghoff, G. von Kiedrowski, *Angew. Chem. Int. Ed.* **47**, 3626 (2008).
47. S. Y. Park *et al.*, *Nature* **451**, 553 (2008).
48. D. Nykypanchuk, M. M. Maye, D. van der Lelie, O. Gang, *Nature* **451**, 549 (2008).
49. J. Liu, D. P. Wernette, Y. Lu, *Angew. Chem. Int. Ed.* **44**, 7290 (2005).
50. F. A. Aldaye, H. F. Sleiman, *J. Am. Chem. Soc.* **129**, 10070 (2007).
51. P. Yin, H. M. T. Choi, C. R. Calvert, N. A. Pierce, *Nature* **451**, 318 (2008).
52. C. Lin, X. Wang, Y. Liu, N. C. Seeman, H. Yan, *J. Am. Chem. Soc.* **129**, 14475 (2007).
53. L. H. Eckardt *et al.*, *Nature* **420**, 286 (2002).
54. X. Li, Z.-Y. Zhan, R. Knipe, D. Lynn, *J. Am. Chem. Soc.* **124**, 746 (2002).
55. J. Bath, A. J. Turberfield, *Nat. Nanotechnol.* **2**, 275 (2008).
56. F.A.A. and H.F.S. have filed a provisional patent, U.S. 60/960,000, on using DNA nanocapsules for controlled drug and gene delivery.

Magnetic Source Separation in Earth's Outer Core

Kenneth A. Hoffman[1,2]* and Brad S. Singer[2]

The dipole component of the geomagnetic field is anomalously strong at both Earth's surface and the core-mantle boundary (CMB). Because dipole terms are of the lowest degree, they are the most capable of reaching the CMB from sources deep within the outer core (1). Nondipole (higher-degree) terms need originate from sources residing closer to the CMB if they are to emerge. Yet, the power associated with the equatorial dipole terms is compatible with the nondipole power spectrum [e.g., (2)], leaving the axial dipole to stand alone given its unique strength. The question then is whether the source of the axial dipole is physically distinct from sources responsible for the rest of the field, the so-called nonaxial dipole (NAD) field.

Analyses of the NAD field at Earth's surface indicate that its present structure is similar to that when time-averaged over the past 400 years (3); the most intense patches of vertical flux (Fig. 1, right) appear almost motionless, indicative of long-term control over shallow core fluid by the lowermost mantle. Hence, the time-averaged NAD field may be used as a proxy for the modern-day NAD field and vice versa.

Twentieth-century observatory data suggest further that these standing features strengthen at independent rates (4). The effect about the globe of this NAD field secular variation is displayed in Fig. 1 (left): Localities close to (and hence dominated by) a single flux feature, versus those more equally proximate to multiple features, experienced lesser and greater vector field changes, respectively.

We focus on two widely separated sites—West Eifel, Germany, and Tahiti, French Polynesia—from which we have available paleomagnetic

transitional field data obtained from lavas that erupted since the Matuyama-Brunhes polarity reversal. The ^{40}Ar/^{39}Ar age determinations of West Eifel lavas indicate the recording of five excursions spanning some 200,000 years, including the Big Lost Event (table S1). The transitional lavas from Tahiti also record the Big Lost Event (3) and the Matuyama-Brunhes reversal (5). Virtual poles recorded in transitionally magnetized lavas from West Eifel are spread across Eurasia, whereas the two events recorded on Tahiti are associated with the same tightly clustered virtual geomagnetic pole (VGP) location west of Australia, where the most intense NAD field flux feature exists at Earth's surface (Fig. 1, right).

Modern-day NAD field structure and behavior tend to explain the paleomagnetic findings: The site of West Eifel lies within the sphere of influence of three concentrations of NAD field flux within the area of greatest directional secular change. In contrast, Tahiti is considerably closer to one, the Australasian feature, and is within the area of least secular change. Thus, both the wide east-west spread in West Eifel Brunhes–aged transitional VGPs and the nearly identical Tahitian VGP clusters are compatible with recent and historic geomagnetic findings.

Also plotted in Fig. 1 (right) are south VGPs associated with the NAD field at both sites throughout the 20th century (6). These recent-field virtual poles, associated with complete removal of the axial dipole term, show behavior similar to the paleomagnetic data, both in angular change and location, for both sites. From these correlations, we

conclude that polarity transitions first involve the demise of the source generating the axial dipole, which leaves the field generated only in the shallow core with a pattern strongly controlled by the physical variability of the lower mantle. Hence, we suggest that there are two significantly independent field sources: one generated deep within the outer core; the other generated in the shallow core, which we designate the SCOR field.

The SCOR field is essentially the NAD field; however, its complex pattern most assuredly contains a small contribution to the axial dipole. The deeper-core field then provides nearly all of the observed axial dipole, yet it must also contain a (small) contribution to lower-degree harmonics in the observed NAD field (7). Such a field source dichotomy may be the key to solving the problem of the reversing geodynamo.

References and Notes

1. Having the longest spatial wavelengths, dipole components are least likely to be screened out by overlying electrically conductive fluid.
2. V. Courtillot, J.-P. Valet, G. Hulot, J.-L. Le Mouel, *Eos Trans. AGU* **73**, 337 (1992).
3. C. G. Constable, in *Encyclopedia of Geomagnetism and Paleomagnetism*, D. G. Gubbins, E. Herrero-Bervera, Eds. (Springer, Dordrecht, Netherlands, 2007), pp. 159–161.
4. K. A. Hoffman, B. S. Singer, *AGU Geophys. Monogr.* **145**, 233 (2004).
5. B. S. Singer *et al.*, *Nature* **434**, 633 (2005).
6. The dynamo is blind to the field sign, so either case is equally valid.
7. Some field asymmetry is required by Cowling's theorem to maintain dynamo action.
8. This study was supported by grants to Cal Poly and the University of Wisconsin-Madison from the Geophysics Program of NSF.

[1]Physics Department, California Polytechnic State University, San Luis Obispo, CA 93407, USA. [2]Department of Geology and Geophysics, University of Wisconsin-Madison, Madison, WI 53706, USA.

*To whom correspondence should be addressed. E-mail: khoffman@calpoly.edu
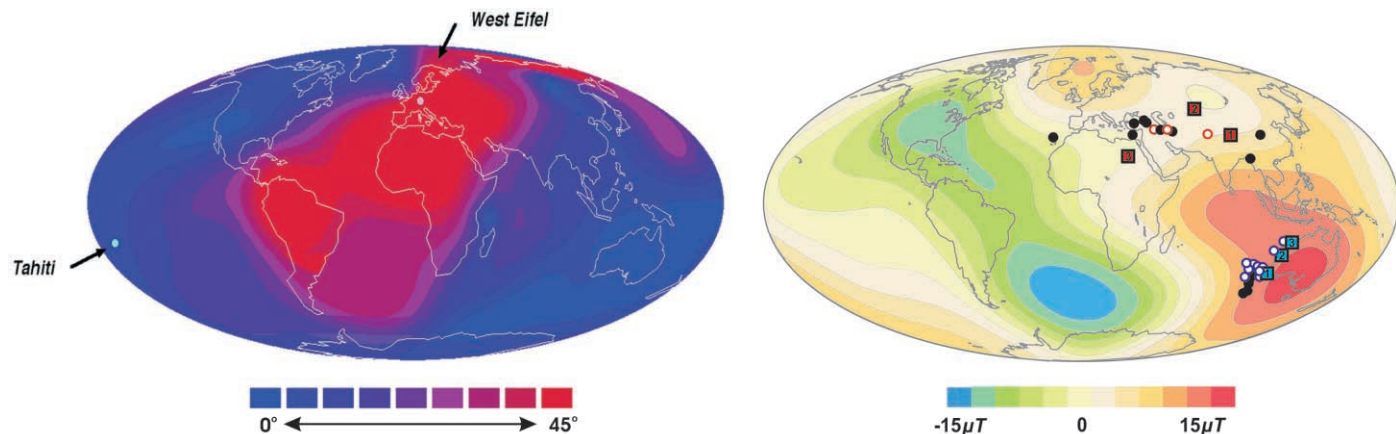
Fig. 1. (**Left**) Contoured angular change from 1900 to 2000 of NAD field VGPs about the globe. (**Right**) (i) Transitional north VGPs recorded in lavas on Tahiti (clustered near west Australia) and West Eifel (spanning much of Eurasia), each case spanning ~200,000 years (Big Lost Event VGPs, which were recorded at both sites, have open symbols), and (ii) concurrent south VGPs for the years 1900, 1950, and 2000 (indicated on the map by 1, 2, and 3, respectively) NAD field at Tahiti (blue squares) and West Eifel (red squares) plotted on the 1590–1990 time-averaged surface NAD field.

Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses

Siân Jones,[1]* Xiaosong Zhang,[1]* D. Williams Parsons,[1,2]* Jimmy Cheng-Ho Lin,[1]*
Rebecca J. Leary,[1]* Philipp Angenendt,[1] Parminder Mankoo,[3] Hannah Carter,[3]
Hirohiko Kamiyama,[4] Antonio Jimeno,[1] Seung-Mo Hong,[4] Baojin Fu,[4] Ming-Tseh Lin,[4]
Eric S. Calhoun,[1] Mihoko Kamiyama,[4] Kimberly Walter,[4] Tatiana Nikolskaya,[5]
Yuri Nikolsky,[6] James Hartigan,[7] Douglas R. Smith,[7] Manuel Hidalgo,[1] Steven D. Leach,[1,8]
Alison P. Klein,[1,4] Elizabeth M. Jaffee,[1,4] Michael Goggins,[1,4] Anirban Maitra,[1,4]
Christine Iacobuzio-Donahue,[1,4] James R. Eshleman,[1,4] Scott E. Kern,[1,4] Ralph H. Hruban,[1,4]
Rachel Karchin,[3] Nickolas Papadopoulos,[1] Giovanni Parmigiani,[1,9] Bert Vogelstein,[1]†
Victor E. Velculescu,[1]† Kenneth W. Kinzler[1]†

There are currently few therapeutic options for patients with pancreatic cancer, and new insights into the pathogenesis of this lethal disease are urgently needed. Toward this end, we performed a comprehensive genetic analysis of 24 pancreatic cancers. We first determined the sequences of 23,219 transcripts, representing 20,661 protein-coding genes, in these samples. Then, we searched for homozygous deletions and amplifications in the tumor DNA by using microarrays containing probes for ~10^6 single-nucleotide polymorphisms. We found that pancreatic cancers contain an average of 63 genetic alterations, the majority of which are point mutations. These alterations defined a core set of 12 cellular signaling pathways and processes that were each genetically altered in 67 to 100% of the tumors. Analysis of these tumors' transcriptomes with next-generation sequencing-by-synthesis technologies provided independent evidence for the importance of these pathways and processes. Our data indicate that genetically altered core pathways and regulatory processes only become evident once the coding regions of the genome are analyzed in depth. Dysregulation of these core pathways and processes through mutation can explain the major features of pancreatic tumorigenesis.

Worldwide, over 213,000 patients will develop pancreatic cancer in 2008, and nearly all will die of their disease (*1–3*). Several genetic alterations have been identified in these lethal cancers, including those in the *CDKN2A*, *SMAD4*, and *TP53* tumor suppressor genes and in the *KRAS* oncogene (*4–8*). Although the discoveries of these genes have provided important insights into the natural history of the disease and have spurred efforts to develop improved diagnostic and therapeutic agents, the vast majority of human genes have not been analyzed in this cancer type.

We examined the genetic makeup of human pancreatic cancers in unprecedented detail. Because all human cancers are primarily genetic diseases, we hoped to identify additional genes and signaling pathways that could guide future research on this disease.

Sequencing strategy. The sequences of protein-coding exons from 20,735 genes were identified and used to design primers for 219,229 amplicons covering these regions (*9*). DNA from 24 advanced pancreatic adenocarcinomas (table S1) was polymerase chain reaction (PCR)–amplified with these primers and sequenced with the use of fluorescent dye terminators (*9*). The 24 tumors were passaged in vitro as cell lines or in nude mice as xenografts to remove contaminating non-neoplastic cells, facilitating detection of mutations (*10–12*). Exons containing variant sequences were reamplified and resequenced from the tumor DNA as well as from normal DNA from the same patient to confirm the mutation and to ensure that the mutation was somatic (i.e., that it was not present in normal cells). PCR products from 208,311 amplicons resulted in PCR products that were successfully sequenced and met stringent quality controls (table S2). These amplicons included 94.5% of the targeted sequences and yielded high-quality sequencing data for 98.5% of the target bases within these amplicons. The 208,311 successfully sequenced amplicons yielded mutational data on 23,219 transcripts representing 20,661 genes.

Somatic mutations. Among the 1562 somatic mutations detected with this strategy, 25.5% were synonymous, 62.4% were missense, 3.8% were nonsense, 5.0% were small insertions and deletions, and 3.3% were at splice sites or within the untranslated region (UTR) (Table 1 and table S3). The spectra of somatic mutations can yield insights into potential carcinogens and other environmental exposures. Table 1 lists the spectra observed in the four tumors that have been subjected to large-scale sequencing analyses of the majority of protein-encoding genes. It is evident that breast tumors have a unique somatic mutation spectrum, with a preponderance of mutations at 5′-TpC sites and a relatively small number of mutations at 5′-CpG sites. However, the spectra of colorectal (*13*, *14*), brain (*15*), and pancreatic tumors are similar, suggesting that breast epithelial cells are exposed to different levels or types of carcinogens or use distinctive repair systems (*16*, *17*). Given that cells in the colon are expected to be exposed to dietary carcinogens more than breast, brain, or pancreatic cells, one possible interpretation of these results is that dietary components are not directly responsible for causing most of the mutations found in human cancers.

Of the 20,661 genes analyzed by sequencing, 1327 had at least one mutation, and 148 had two or more mutations among the 24 cancers surveyed (table S3). In addition to the frequency of mutations, the type of mutation can provide information useful for evaluating its potential role in disease (*18*). Nonsense mutations, out-of-frame insertions or deletions, and splice-site changes generally lead to inactivation of the protein products. To evaluate missense mutations, we developed an algorithm that uses machine learning of 58 predictive features based on the physical-chemical properties of amino acids involved in the substitutions and their evolutionary conservation at equivalent positions of conserved proteins (*9*). Of the 924 missense mutations that could be scored with this algorithm, 160 (17.3%) were predicted to contribute to tumorigenesis when assessed by this method (table S3).

We also generated structural models of 404 of the missense mutations identified in this study [links to structural models available at (*19*)]. In each case, the model was based on x-ray crystallography or nuclear magnetic resonance spectroscopy of the normal protein or a closely related homolog. This analysis showed that 55 of the 404 mutations were located near a domain interface or ligand-binding site and were likely to affect function (examples in Fig. 1).

The average number of somatic mutations in pancreatic cancers (48; Table 2) is considerably

[1]Sol Goldman Pancreatic Cancer Research Center, Ludwig Center and Howard Hughes Medical Institute at the Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. [2]Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston, TX 77030, USA. [3]Department of Biomedical Engineering, Institute of Computational Medicine, Johns Hopkins Medical Institutions, Baltimore, MD 21218, USA. [4]Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. [5]Vavilov Institute for General Genetics, Moscow B333, 117809, Russia. [6]GeneGo, Incorporated, St. Joseph, MI 49085, USA. [7]Agencourt Bioscience Corporation, Beverly, MA 01915, USA. [8]Department of Surgery, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. [9]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: bertvog@gmail.com (B.V.); velculescu@jhmi.edu (V.E.V.); kinzlke@jhmi.edu (K.W.K.)

less than that in breast cancer (101) or colorectal cancers (77) ($P < 0.001$), even though fewer genes were sequenced in the latter two tumor types (14). One plausible explanation for this lower rate is that the cells that initiate pancreatic tumorigenesis have gone through fewer divisions than colorectal or breast cancer cells. It has been previously shown that the majority of mutations observed in colorectal cancers are likely to have occurred in the normal stem cells that gave rise to the initiating neoplastic cell (12). Our data are thus consistent with observations showing that normal pancreatic epithelial cells divide infrequently (20, 21).

We further evaluated 39 genes that were mutated in more than one of the 24 discovery screen cancers in a prevalence screen consisting of 90 pancreatic cancers. In this screen, we detected 255 nonsilent somatic mutations among 23 genes (table S4). The nonsilent mutation rate of the genes in the prevalence screen (excluding KRAS, TP53, CDK2NA, and SMAD4) was higher than that in the discovery screen (3.6 versus 1.47 nonsilent mutations per Mbase, $P < 0.001$). The fraction of nonsilent mutations observed in these 19 genes was also higher than that observed in the genes assessed in the discovery screen ($P = 0.052$). These data are consistent with the hypothesis that a greater fraction of the genes tested in the prevalence screen were positively selected during tumorigenesis.

Deletions. By using oligonucleotide arrays containing probes for 1,069,688 SNPs and robust algorithms for identifying deletion events from SNP array data (22), we identified 198 separate homozygous deletions among the 24 pancreatic cancers (table S5). The average size of these deletions was 335,000 bp. Additionally, we observed many regions that had undergone single-copy losses, often manifest as losses of heterozygosity, including losses of whole chromosomes or whole chromosome arms. We did not pursue these changes because it is difficult to reliably identify target genes from such large regions unless the residual copy of the gene on the nondeleted chromosome is mutated. Such target genes would have already been called to our attention by the results of the discovery sequencing screen and would have been scored as homozygous changes (table S3).

According to the allelic two-hit hypothesis, the presence of a homozygous deletion indicates that a tumor suppressor gene exists within the deleted region (23). To determine the most likely target within these deletions, we used the results from our analysis of point mutations as well as expression analyses (see below) and previously published studies. For a gene to be considered the target, a portion of its coding region had to be affected by the homozygous deletion, and the gene (i) had to harbor a nonsilent sequence alteration in a different tumor, (ii) had to be a well-documented tumor suppressor gene, or (iii) had to have corroborating expression data. The presumptive target genes for the homozygous

deletions that met these criteria are listed in table S5. This list includes the classic tumor suppressor genes CDKN2A (p16), SMAD4, and TP53, as well as genes that had not previously been implicated in pancreatic cancer development.

When an exon of a gene is truly deleted in a tumor, no sequencing information should be obtainable, providing confirmation of the deletion. Without exception, the homozygous deletions found through the SNP arrays were consistent with the sequencing data (9). Furthermore, there was only one homozygous deletion revealed by sequencing that was not evident in the microarray hybridizations (a four-exon deletion of SMAD4 in tumor Pa21C).

The number of deletions in a tumor was more variable than the number of somatic mutations, averaging 8.3 and ranging between 2 and 20 per tumor (Fig. 2). However, each homozygous deletion completely abrogated the function of the target gene as well as all other genes within the deleted region, whereas only a fraction of the somatic mutations were predicted to alter the gene's function. In a typical pancreatic cancer, ~10 genes (including targets and nearby genes) are eradicated by homozygous deletion, providing fertile grounds for therapeutic strategies that target such losses (24, 25).

Amplifications. With the use of algorithms similar to those described above for deletions (22), we identified 144 focal high-copy amplifi-

cations in the 24 tumors (table S6). We also identified a variety of low-copy-number gains of entire chromosomes, chromosomal arms, or other large genomic regions that were not pursued because of the difficulty in reliably identifying candidate cancer genes from such large chromosomal regions. To determine the most likely target of the focal amplifications, we again used the results from our mutational data, expression analyses, and previously published data. The presumptive target genes for each of the amplifications that met predefined criteria (9) are listed in table S6. There were fewer amplifications than homozygous deletions or point mutations in most pancreatic tumors (Fig. 2).

Passenger mutation rates. The primary goal of cancer genome studies is the identification of genes that are likely to play a causal role in the neoplastic process (potential drivers). One can categorize the best candidate cancer genes (CAN genes) on the basis of their mutation frequencies and types. This categorization requires an estimate of the passenger mutation rate (13, 14, 26). For each of the genes containing somatic mutations, passenger probabilities were determined by using estimated minimal and maximal passenger mutation rates after taking into account the size of the gene, its nucleotide composition, and other relevant factors [Table 1 and (9)]. To analyze the probability that a given gene would be involved in an amplification or deletion, we made the

Table 1. Summary of somatic mutations in four tumor types. Pancreas data have their basis in 24 tumors analyzed in the current study; brain data have their basis in 21 nonhypermutable tumors analyzed in (15); and colorectal and breast data have their basis in 11 breast and 11 colorectal tumors analyzed in (14). Nonsilent numbers in parentheses refer to percentage of total nonsynonymous mutations, and substitutions numbers in parentheses refer to percentage of total substitutions. The total number of substitutions includes synonymous as well as nonsilent mutations identified in the indicated study.

	Pancreas	Brain	Colorectal	Breast
Number of mutated genes	1007	685	769	1026
Number of nonsilent mutations	1163	748	849	1112
Missense	974 (83.7)	622 (83.2)	722 (85)	909 (81.7)
Nonsense	60 (5.2)	43 (5.7)	48 (5.7)	64 (5.8)
Insertion	4 (0.3)	3 (0.4)	4 (0.5)	5 (0.4)
Deletion	43 (3.7)	46 (6.1)	27 (3.2)	78 (7.0)
Duplication	31 (2.7)	7 (0.9)	18 (2.1)	3 (0.3)
Splice site or UTR	51 (4.4)	27 (3.6)	30 (3.5)	53 (4.8)
Total number of substitutions	1484	937	893	1157
Substitutions at C:G base pairs				
C:G to T:A	798 (53.8)	601 (64.1)	534 (59.8)	422 (36.5)
C:G to G:C	142 (9.6)	67 (7.2)	61 (6.8)	325 (28.1)
C:G to A:T	246 (16.6)	114 (12.1)	130 (14.6)	175 (15.1)
Substitutions at T:A base pairs				
T:A to C:G	142 (9.6)	87 (9.3)	69 (7.7)	102 (8.8)
T:A to G:C	79 (5.3)	24 (2.6)	59 (6.6)	57 (4.9)
T:A to A:T	77 (5.2)	44 (4.7)	40 (4.5)	76 (6.6)
Substitutions at specific dinucleotides				
5'-CpG-3'	563 (37.9)	404 (43.1)	427 (47.8)	195 (16.9)
5'-TpC-3'	218 (14.7)	102 (10.9)	99 (11.1)	395 (34.1)

conservative assumption that the overall frequency of all observed amplifications and deletions represented the passenger mutation rate (9).

Passenger probabilities for all genes in which at least two genetic alterations were identified in the discovery screen are listed in table S7. This list includes all genes previously known to play an important role in pancreatic cancer through mutation or copy number change, providing experimental confirmation of our general approach. Importantly, the *CAN* genes listed in table S7

included numerous other genes of potential biological interest, many of which had not previously been identified as playing an important role in this tumor type. Examples include the transcriptional activator *MLL3*; cadherin homologs *CDH10*, *PCDH15*, and *PCDH18*; the α-catenin *CTNNA2*; the dipeptidyl-peptidase *DPP6*; the angiogenesis inhibitor *BAI3*; the heterotrimeric guanine nucleotide–binding protein (G-protein)–coupled receptor *GPR133*; the guanylate cyclase *GUCY1A2*; the protein kinase *PRKCG*; and

Q9H5F0, a gene of unknown function. These genes were generally mutated at much lower frequencies than those previously identified to be mutated in pancreatic cancers (table S7). This is compatible with the idea that conventional strategies were able to identify frequently mutated genes but not the bulk of the genes that are genetically altered in pancreatic cancers.

Candidate pathways and processes promoting pancreatic tumorigenesis. Because most cellular pathways and processes involve multiple
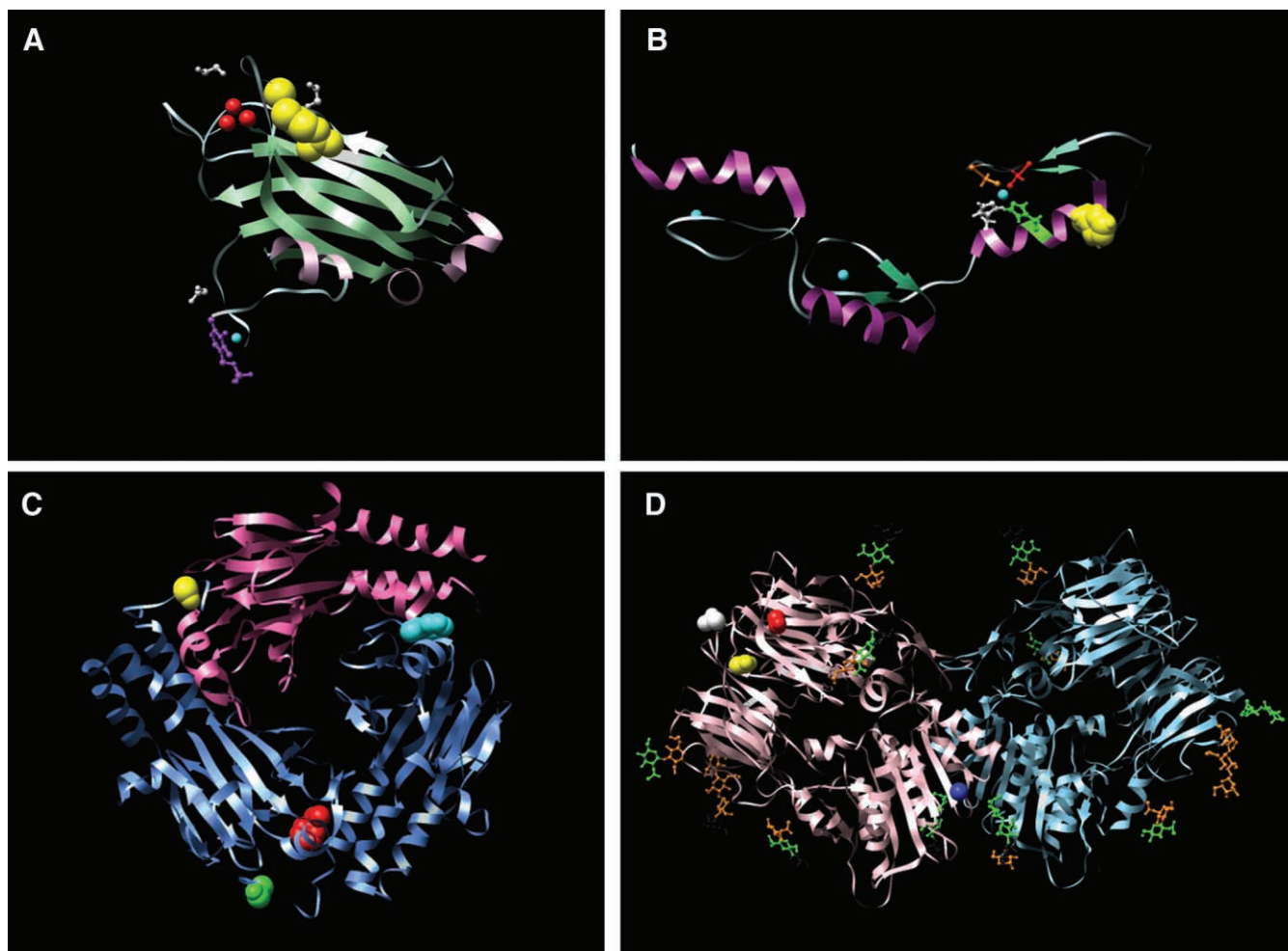
Fig. 1. Examples of structural models of mutations. (**A**). The x-ray crystal structure of the C2 domain of protein kinase C γ (PKCG) [Protein Data Bank identification number (PDBID) 2UZP]. R252 (41) is shown as yellow space-fills; Ca^{2+} ions are shown as red spheres. The ligands 1,2-ethanediol and pyridoxal-5-phosphate are shown in white and purple ball-and-stick representations, respectively. The R252→H252 (R252H) mutation could reduce the membrane binding of the C2 domain of PRKCG and thereby affect function. (**B**) The nuclear magnetic resonance solution structure of the three tandem repeats of zf-C2H2 domains from human Kruppel-like factor 5 (KLF5) (PDBID 2EBT). H389 is shown as yellow space-fills; Zn^{2+} ions are shown as cyan spheres. The residues comprising the C2H2 group that coordinate the nearby Zn^{2+} ion are shown as ball-and-stick representations, H393 and H397 are shown in green and white, whereas C380 and C375 are shown in orange and red. The mutation at position 389 (H389N) may disrupt the structure of the zinc finger or nearby zinc coordination site. (**C**) The x-ray crystal structure of the heterotrimer of SMAD3 (two subunits shown as blue ribbons) and SMAD4 (one subunit shown as pink ribbons) (PDBID 1U7F). The residues corresponding to two of the mutant positions (F260S and S422F, shown as red and yellow space-fills, respectively, in chain A) are located at interfaces and could perturb Smad3-Smad3 or Smad3-Smad4 interactions. In chain B, F260 is shown as cyan space-fills and S422 as green space-fills. (**D**) The x-ray crystal structure of the extracellular domain of human DPP6 as a homodimer (PDBID 1XFD). Two of the mutated residues found in this study, T409I (shown as red space-fills) and D475N (shown in yellow space-fills) are in spatial proximity and are close to one of the glycosylation sites, N471 (shown as white space-fills). These mutations fall in the β-propeller domain of the protein (residues 142 to 322 and 351 to 581) thought to be involved in protein-protein interactions. The A778T mutation (shown as blue space-fills) falls in the */β hydrolase domain (residues 127 to 142 and 581 to 849) and is close to the homodimer region of the protein and could perturb the homodimer association. Carbohydrates with glycosylation sites are shown in stick representation. Images created with UCSF Chimera version 1.2422 for Linux (42).

proteins functioning in a concerted manner, it is possible that mutations in different genes result in similar tumorigenic effects. Because nearly all of the protein-coding genes in the human genome were evaluated in the current study, the data provided a unique opportunity to investigate groups of genes operating though specific signaling pathways and processes. Sets of genes involved in signaling pathways or cellular processes were defined through three well-annotated GeneGo MetaCore databases: gene ontology (GO), canonical gene pathway maps (MA), and genes participating in defined cellular processes and networks (GG) (27). We developed a statistical approach that provided a combined probability that a gene set contained driver alterations, taking into account all types of genetic alterations evaluated in this study (22). For each gene set, we considered whether the component genes were more likely to be affected by a genetic alteration than would be predicted by the passenger mutation rate.

These analyses identified 69 gene sets that were genetically altered in the majority of the 24 cancers examined (table S8). Thirty-one of these sets could be further grouped into 12 core signaling pathways and processes that were each altered in 67 to 100% of the 24 cancers analyzed and had clear functional relevance to neoplasia based on annotations in the databases described above (Table 2). The core pathways included those in which a single, frequently altered gene predominated, such as in KRAS signaling and in the regulation of the G1/S cell cycle transition; pathways in which a few altered genes predominated, such as in TGF-β signaling; and pathways in which many different genes were altered, such as in integrin signaling, regulation of invasion,

homophilic cell adhesion, and small guanine triphosphatase (GTPase)–dependent signaling.

Analysis of gene expression. Gene expression patterns can inform the analysis of pathways because they can reflect epigenetic alterations not detectable by sequencing or copy number analyses. They can also point to downstream effects on gene expression resulting from the altered signaling pathways and processes described above. To analyze the transcriptome of pancreatic cancers, we performed SAGE [serial analysis of gene expression (28)] on RNA from the same 24 cancers used for mutation analysis. When combined with massively parallel sequencing by synthesis, SAGE provides a highly quantitative and sensitive measure of gene expression. The approach described above is similar to that used in recent RNA-Seq studies (29–32), but SAGE has the advantage that the quantification does not depend on the length of the transcript.

As a control for the current study, we microdissected histologically normal pancreatic duct epithelial cells because these cells are the presumed precursors of pancreatic cancers. As an additional control, we used human papillomavirus (HPV)–immortalized pancreatic duct epithelial (HPDE) cells, which have been shown to have many properties in common with normal duct epithelial cells (33, 34). SAGE libraries were prepared from these cells as well as the 24 pancreatic cancers; an average of 5,737,000 tags was obtained from each library, and an average of 2,268,000 tags per library matched the sequence of known transcripts (table S9).

The expression analysis was first used to help identify target genes from amplified and homozygously deleted regions. Although a small fraction of

these regions contained a known tumor suppressor gene or oncogene, many contained more than one gene that had not previously been implicated in cancer. In tables S5 and S6, a presumptive target gene was identified within these regions through the use of the mutational and transcriptional data. For example, we assumed that a gene could not have been the target of an amplification event if that gene was not wholly contained within the amplicon and expressed in the tumor containing the amplification. Similarly, expression data can be used to help gauge the importance of genes containing missense mutations. A missense mutation in a gene that is not expressed in the tumor containing it is more likely to be a passenger than a mutation in a gene that is expressed (table S3).

Second, we determined whether the genes in the core signaling pathways and processes described above were differentially expressed. If the pathways and processes containing genetic alterations were indeed responsible for tumorigenesis, one might expect that many of the genes within these pathways would be aberrantly expressed. To test this hypothesis, we examined the expression of the gene sets constituting the 12 core signaling pathways and processes (Table 2 and table S8). The 31 gene sets constituting these pathways were more highly enriched for differentially expressed genes than the remaining 3041 gene sets ($P <$ 0.001). These expression data thus independently support the contribution of these signaling pathways and processes to pancreatic tumorigenesis.

Lastly, we attempted to identify individual genes rather than pathways that were differentially expressed in the cancers. The data in table S9 represent the largest compendium of digital expression data derived for any tumor type to
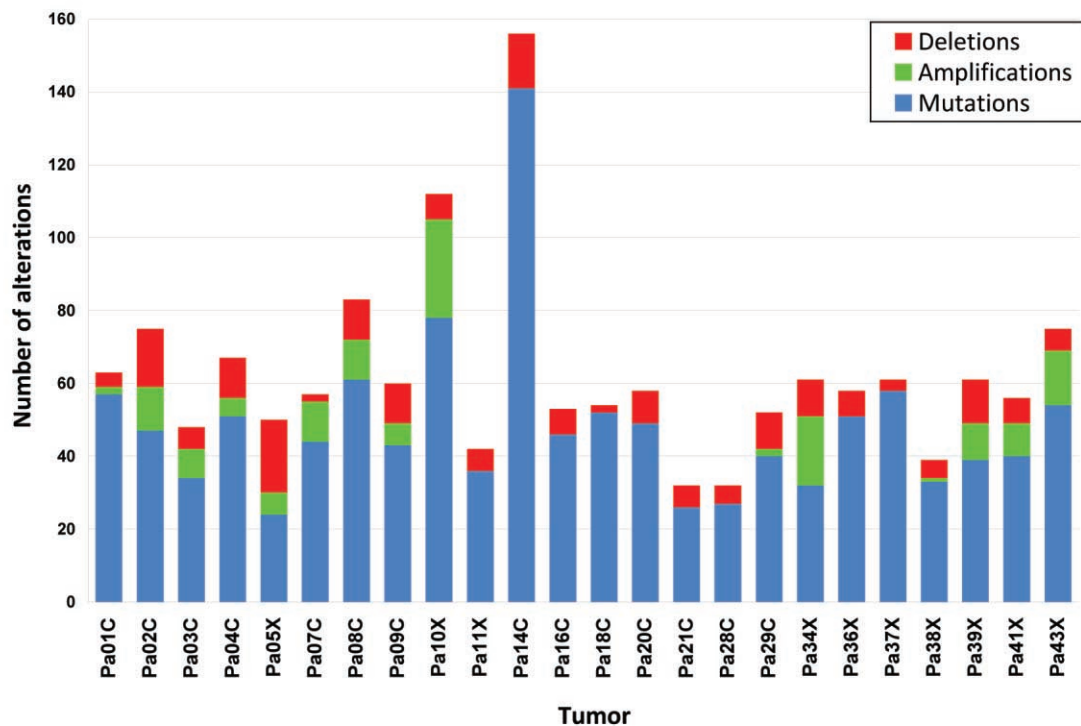
Fig. 2. Number of genetic alterations detected through sequencing and copy number analyses in each of the 24 cancers.

date. There was a remarkably high number (541) of genes that were at least 10-fold overexpressed in >90% of the 24 cancers (compared to normal pancreatic duct cells or HPDE cells). To determine whether these genes were also overexpressed in the primary tumors from which the cell lines were made, we performed SAGE on five such primary tumors. These results confirmed these 541 genes' overexpression in situ: The genes were, on average, expressed at 75-fold higher levels in the cell lines and at 88-fold higher levels in the primary tumors compared with their expression in normal duct epithelial cells. Notably, 54 of the overexpressed genes encoded proteins that are predicted to be secreted or expressed on the cell surface (table S9). These overexpressed genes provide leads for a variety of diagnostic and therapeutic approaches.

Implications for Pancreatic Tumorigenesis

The extensive genetic studies described above suggest that the key to understanding pancreatic cancers lies in an appreciation of a core set of pathways and processes. We identified 12 partially overlapping processes that are genetically altered in the great majority of pancreatic cancers (Fig. 3A). However, the pathway components that are altered in any individual tumor vary widely (Fig. 3, B and C). For example, the two tumors depicted in Fig. 3, B and C, each contain mutations of a gene involved in the TGF-β pathway (one *SMAD4*, the other *BMPR2*). Similarly, these two tumors both contain mutations of genes involved in most of the other 11 core processes and pathways, but the specific genes altered in each tumor are largely different. Although we cannot be certain that every identified mutation plays a functional role in the pathway or process in which it is implicated, it is clear from both the current and the previously published genetic data, as well as from past functional studies, that many of them are likely to affect these pathway(s).

This perspective is likely to apply to most, if not all, epithelial tumors. It is consistent with the idea that genetic alterations can be classified as mountains (high-frequency mutations) or hills (low-frequency mutations), with the hills predominating in terms of the total number of alterations involved (*14*). The heterogeneity among pathway components and the varied nature of mutations within individual genes can explain tumor heterogeneity, a fundamental facet of all solid tumors (*35*).

From an intellectual viewpoint, the pathway perspective helps bring order and rudimentary understanding to a very complex disease (*36–38*). Although the importance of signaling pathways in understanding neoplasia has been recognized (*39, 40*), genomewide genetic analyses such as that described here can identify the precise genetic alterations that may be responsible for pathway dysregulation in each patient's tumor. Because most genes are mutated in only a small fraction of tumors, it is only through analysis of functional gene groups that an appreciation for the true importance of these genes' mutations in neoplasia can be reached. For example, from Table 2 it

Table 2. Core signaling pathways and processes genetically altered in most pancreatic cancers. A complete listing of the gene sets defining these signaling pathways and processes and the statistical significance of each gene set are provided in table S8.

Regulatory process or pathway	Number of genetically altered genes detected	Fraction of tumors with genetic alteration of at least one of the genes	Representative altered genes
Apoptosis	9	100%	*CASP10, VCP, CAD, HIP1*
DNA damage control	9	83%	*ERCC4, ERCC6, EP300, RANBP2, TP53*
Regulation of G_1/S phase transition	19	100%	*CDKN2A, FBXW7, CHD1, APC2*
Hedgehog signaling	19	100%	*TBX5, SOX3, LRP2, GLI1, GLI3, BOC, BMPR2, CREBBP*
Homophilic cell adhesion	30	79%	*CDH1, CDH10, CDH2, CDH7, FAT, PCDH15, PCDH17, PCDH18, PCDH9, PCDHB16, PCDHB2, PCDHGA1, PCDHGA11, PCDHGC4*
Integrin signaling	24	67%	*ITGA4, ITGA9, ITGA11, LAMA1, LAMA4, LAMA5, FN1, ILK*
c-Jun N-terminal kinase signaling	9	96%	*MAP4K3, TNF, ATF2, NFATC3*
KRAS signaling	5	100%	*KRAS, MAP2K4, RASGRP3*
Regulation of invasion	46	92%	*ADAM11, ADAM12, ADAM19, ADAM5220, ADAMTS15, DPP6, MEP1A, PCSK6, APG4A, PRSS23*
Small GTPase–dependent signaling (other than KRAS)	33	79%	*AGHGEF7, ARHGEF9, CDC42BPA, DEPDC2, PLCB3, PLCB4, RP1, PLXNB1, PRKCG*
TGF-β signaling	37	100%	*TGFBR2, BMPR2, SMAD4, SMAD3*
Wnt/Notch signaling	29	100%	*MYC, PPP2R3A, WNT9A, MAP2,TSC2, GATA6, TCF4*
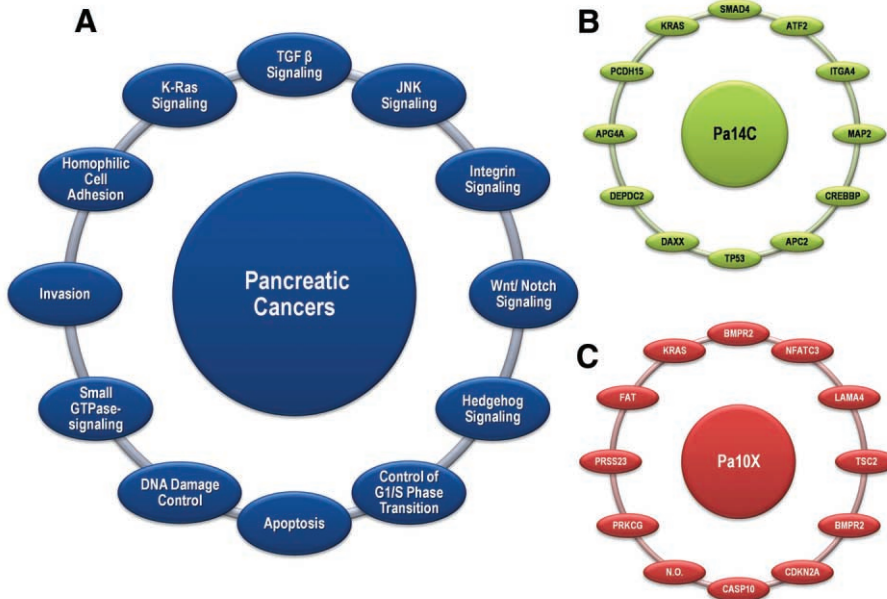
Fig. 3. Signaling pathways and processes. (**A**) The 12 pathways and processes whose component genes were genetically altered in most pancreatic cancers. (**B** and **C**) Two pancreatic cancers (Pa14C and Pa10X) and the specific genes that are mutated in them. The positions around the circles in (B) and (C) correspond to the pathways and processes in (A). Several pathway components overlapped, as illustrated by the BMPR2 mutation that presumably disrupted both the SMAD4 and Hedgehog signaling pathways in Pa10X. Additionally, not all 12 processes and pathways were altered in every pancreatic cancer, as exemplified by the fact that no mutations known to affect DNA damage control were observed in Pa10X. N.O. indicates not observed.

is evident that all pancreatic cancers studied had alterations in genes in the Wnt/Notch and Hedgehog signaling pathways, a finding that could not have been appreciated in the absence of global analyses.

In addition to yielding insights into tumor pathogenesis, such studies provide the data required for personalized cancer medicine. Unlike certain forms of leukemia, in which tumorigenesis appears to be driven by a single, targetable oncogene, pancreatic cancers result from genetic alterations of a large number of genes that function through a relatively small number of pathways and processes. Our studies suggest that the best hope for therapeutic development may lie in the discovery of agents that target the physiologic effects of the altered pathways and processes rather than their individual gene components. Thus, rather than seeking agents that target specific mutated genes, agents that broadly target downstream mediators or key nodal points may be preferable. Pathways that could be targeted include those causing metabolic disturbances, neoangiogenesis, misexpression of cell surface proteins, alterations of the cell cycle, cytoskeletal abnormalities, and an impaired ability to repair genomic damage (Table 2 and table S8).

References and Notes

1. D. M. Parkin, F. I. Bray, S. S. Devesa, *Eur. J. Cancer* **37** (suppl.), S4 (2001).
2. A. Jemal *et al.*, *CA Cancer J. Clin.* **58**, 71 (2008).
3. J. B. Koorstra, S. R. Hustinx, G. J. Offerhaus, A. Maitra, *Pancreatology* **8**, 110 (2008).
4. E. Efthimiou, T. Crnogorac-Jurcevic, N. R. Lemoine, *Pancreatology* **1**, 571 (2001).
5. M. Mimeault, R. E. Brand, A. A. Sasson, S. K. Batra, *Pancreas* **31**, 301 (2005).
6. D. A. Tuveson, S. R. Hingorani, *Cold Spring Harb. Symp. Quant. Biol.* **70**, 65 (2005).
7. E. M. Jaffee, R. H. Hruban, M. Canto, S. E. Kern, *Cancer Cell* **2**, 25 (2002).
8. A. Maitra, R. H. Hruban, *Annu. Rev. Pathol.* **3**, 157 (2008).
9. Materials and methods are available as supporting material on *Science* Online.
10. J. M. Winter, J. R. Brody, S. E. Kern, *Cancer Biol. Ther.* **5**, 360 (2006).
11. B. Rubio-Viqueira *et al.*, *Clin. Cancer Res.* **12**, 4652 (2006).
12. S. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4283 (2008).
13. T. Sjoblom *et al.*, *Science* **314**, 268 (2006).
14. L. D. Wood *et al.*, *Science* **318**, 1108 (2007).
15. D. W. Parsons *et al.*, *Science* **321**, 1807 (2008).
16. A. Hartmann, H. Blaszyk, J. S. Kovach, S. S. Sommer, *Trends Genet.* **13**, 27 (1997).
17. S. P. Hussain, C. C. Harris, *Mutat. Res.* **428**, 23 (1999).
18. P. C. Ng, S. Henikoff, *Nucleic Acids Res.* **31**, 3812 (2003).
19. R. Karchin, Structural models of mutants identified in pancreatic cancers, http://karchinlab.org/Mutants/CAN-genes/pancreatic/Pancreatic_cancer.html (2008).
20. W. M. Klein, R. H. Hruban, A. J. Klein-Szanto, R. E. Wilentz, *Mod. Pathol.* **15**, 441 (2002).
21. H.-P. Elsasser, G. Adler, H. F. Kern, in *The Pancreas*, V. L. W. Go *et al.*, Eds. (Raven, New York, 1993), pp. 75–86.
22. R. J. Leary *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
23. A. G. Knudson, *Am. J. Med. Genet.* **111**, 96 (2002).
24. S. R. Hustinx *et al.*, *Mod. Pathol.* **18**, 959 (2005).
25. A. Varshavsky, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14935 (2007).
26. C. Greenman, R. Wooster, P. A. Futreal, M. R. Stratton, D. F. Easton, *Genetics* **173**, 2187 (2006).
27. S. Ekins, Y. Nikolsky, A. Bugrim, E. Kirillov, T. Nikolskaya, *Methods Mol. Biol.* **356**, 319 (2007).
28. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995).
29. M. Sultan *et al.*, *Science* **321**, 956 (2008).
30. R. Lister *et al.*, *Cell* **133**, 523 (2008).
31. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nat. Methods* **5**, 621 (2008).
32. R. Morin *et al.*, *Biotechniques* **45**, 81 (2008).
33. T. Furukawa *et al.*, *Am. J. Pathol.* **148**, 1763 (1996).
34. H. Ouyang *et al.*, *Am. J. Pathol.* **157**, 1623 (2000).
35. A. H. Owens, D. S. Coffey, S. B. Baylin, *Tumor Cell Heterogeneity* (Academic Press, New York, 1982).
36. J. Lin *et al.*, *Genome Res.* **17**, 1304 (2007).
37. T. W. Chittenden *et al.*, *Genomics* **91**, 508 (2008).
38. E. J. Edelman, J. Guinney, J. Chi, P. Febbo, S. Mukherjee, *PLOS Comput. Biol.* **4**, e28 (2008).
39. D. Hanahan, R. A. Weinberg, *Cell* **100**, 57 (2000).
40. B. Vogelstein, K. W. Kinzler, *Nat. Med.* **10**, 789 (2004).
41. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
42. E. F. Pettersen *et al.*, *J. Comput. Chem.* **25**, 1605 (2004).
43. Contact information for the authors who directed the major components of this project is as follows: J. R. Eshleman, S. E. Kern (clinical and sample coordination), jeshlem@jhmi.edu, sk@mail.jhmi.edu; R. H. Hruban (pathological review), rhruban@jhmi.edu; R. Karchin (bioinformatic analysis), karchin@jhu.edu; N. Papadopoulos (gene expression analysis), npapado1@jhmi.edu; G. Parmigiani (statistical analysis), gp@jhu.edu; B. Vogelstein, V. E. Velculescu, K. W. Kinzler (sequencing and copy number analysis), bertvog@gmail.com, velculescu@jhmi.edu, kinzlke@jhmi.edu. We thank J. Ptak, N. Silliman, L. Dobbyn, M. Whalen, M. Borges, G. Cusatis, M. Griffith, C. Henderson, C. Karikari, G. Mo, M. Mullendore, E. Palmisano, M. Raben, S. Solt, and D. Trusty for expert technical assistance; T. Sjöblom for help with database management for mutational analyses; Y. Ding for help with bioinformatic analyses; and J. Cameron, G. Feldmann, S. Salaria, L. Hua, and C. Yeo for helpful discussions. This project was performed under the auspices of the Goldman Pancreatic Cancer Genome Initiative and was funded by the Sol Goldman Charitable Trust; the Lillian Goldman Charitable Trust; the Lustgarten Foundation for Pancreatic Cancer Research; the Virginia and D. K. Ludwig Fund for Cancer Research; the Susan G. Komen Foundation; the Michael Rolfe Pancreatic Cancer Foundation; the Joseph C. Monastra Foundation for Pancreatic Cancer; the family and friends of George Rubis; the Viragh Family Foundation; the Broad Foundation; the Emerald Foundation; NIH grants CA62924, CA43460, CA57345, and CA121113; and Beckman Coulter Corporation. Under separate licensing agreements between the Johns Hopkins University and Genzyme, Beckman Coulter, and Exact Sciences Corporations, K.W.K., B.V., and V.E.V. are entitled to a share of royalties received by the university on sales of products related to research described in this paper. These authors and the university own Genzyme and Exact Sciences stock, which is subject to certain restrictions under university policy. The terms of these arrangements are managed by the Johns Hopkins University in accordance with its conflict-of-interest policies.

An Integrated Genomic Analysis of Human Glioblastoma Multiforme

D. Williams Parsons, *et al.*
Science **321**, 1807 (2008);
DOI: 10.1126/science.1164382

AAAS

The following resources related to this article are available online at
www.sciencemag.org (this information is current as of September 28, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/321/5897/1807

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/1164382/DC1

This article **cites 47 articles**, 27 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/321/5897/1807#otherarticles

This article appears in the following **subject collections**:
Medicine, Diseases
http://www.sciencemag.org/cgi/collection/medicine

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl

An Integrated Genomic Analysis of Human Glioblastoma Multiforme

D. Williams Parsons,[1,2]* Siân Jones,[1]* Xiaosong Zhang,[1]* Jimmy Cheng-Ho Lin,[1]*
Rebecca J. Leary,[1]* Philipp Angenendt,[1]* Parminder Mankoo,[3] Hannah Carter,[3] I-Mei Siu,[4]
Gary L. Gallia,[4] Alessandro Olivi,[4] Roger McLendon,[5] B. Ahmed Rasheed,[5] Stephen Keir,[5]
Tatiana Nikolskaya,[6] Yuri Nikolsky,[7] Dana A. Busam,[8] Hanna Tekleab,[8] Luis A. Diaz Jr.,[1]
James Hartigan,[9] Doug R. Smith,[9] Robert L. Strausberg,[8] Suely Kazue Nagahashi Marie,[10]
Sueli Mieko Oba Shinjo,[10] Hai Yan,[5] Gregory J. Riggins,[4] Darell D. Bigner,[5]
Rachel Karchin,[3] Nick Papadopoulos,[1] Giovanni Parmigiani,[1] Bert Vogelstein,[1]†
Victor E. Velculescu,[1]† Kenneth W. Kinzler[1]†

Glioblastoma multiforme (GBM) is the most common and lethal type of brain cancer. To identify
the genetic alterations in GBMs, we sequenced 20,661 protein coding genes, determined the
presence of amplifications and deletions using high-density oligonucleotide arrays, and performed
gene expression analyses using next-generation sequencing technologies in 22 human tumor
samples. This comprehensive analysis led to the discovery of a variety of genes that were not known
to be altered in GBMs. Most notably, we found recurrent mutations in the active site of isocitrate
dehydrogenase 1 (*IDH1*) in 12% of GBM patients. Mutations in *IDH1* occurred in a large fraction of
young patients and in most patients with secondary GBMs and were associated with an increase in
overall survival. These studies demonstrate the value of unbiased genomic analyses in the
characterization of human brain cancer and identify a potentially useful genetic alteration for the
classification and targeted therapy of GBMs.

Malignant gliomas are the most frequent and lethal cancers originating in the central nervous system. The most biologically aggressive subtype is glioblastoma multiforme (GBM) [World Health Organization (WHO) grade IV astrocytoma], a tumor associated with a dismal prognosis (*1*). The current standard of care for GBM patients—surgical resection followed by adjuvant radiation therapy and chemotherapy with the oral alkylating agent temozolomide—produces a median survival of only 15 months (*2*). Historically, GBMs have been categorized into two groups ("primary" and "secondary") on the basis of clinical presentation (*3*). Secondary GBMs are defined as cancers that have clinical, radiologic, or histopathologic evidence of malignant progression from a preexisting lower-grade tumor, whereas primary GBMs have no such history and present at diagnosis as advanced cancers (*4*). Clinical differences have been reported between the two groups, with secondary GBMs occurring less frequently (~5% of GBMs) and predominantly in younger patients (median age ~45 years versus ~60 years for primary GBM) (*5, 6*). The histopathologic findings of primary and secondary GBMs are indistinguishable, and the prognosis does not appear to be different after adjustment for age (*5, 6*).

Substantial research effort has focused on the identification of genetic alterations in GBMs that might help define subclasses of GBM patients with differing prognoses and/or response to specific therapies (*7*). Distinctions between the genetic lesions found in primary and secondary GBMs have been made, with *TP53* mutations occurring more commonly in secondary GBMs and *EGFR* amplifications and *PTEN* mutations occurring more frequently in primary GBMs (*6, 8, 9*); however, none of these alterations is sufficiently specific to distinguish between primary and secondary GBMs. This issue is further confounded by the possibility that a fraction of GBMs designated as primary tumors may follow a sequence of genetic events similar to that of secondary lesions but not come to clinical attention until malignant progression to a GBM has occurred.

The comprehensive elucidation of genetic alterations in GBMs could provide novel targets that might be used for diagnostic, prognostic, or therapeutic purposes as well as to identify subgroups of patients that preferentially respond to particular targeted therapies. The determination of the human genome sequence and improvements in sequencing and bioinformatic technologies have recently permitted genome-wide sequence analyses in human cancers. We have previously studied the genomes of 11 breast and 11 colorectal cancers by determining the sequence of the more than 18,000 Consensus Coding Sequence (CCDS) and Reference Sequence (RefSeq) genes (*10, 11*). Here, we have analyzed 20,661 protein coding genes in 22 human GBM samples. To complement these sequencing data, we have also performed a genome-wide analysis of focal copy number alterations, including amplifications and homozygous deletions, using high-density oligonucleotide microarrays on the same GBM tumors. Finally, we have examined the expression profiles of these same samples using serial analysis of gene expression (SAGE) and next-generation sequencing technologies.

Sequencing strategy. We extended our previous sequencing strategy for identification of somatic mutations to include 23,219 transcripts from 20,661 genes (fig. S1). These included 2783 additional genes from the Ensembl databases that were not present in the CCDS or RefSeq databases analyzed in the previous studies (*10, 11*). In addition, we redesigned polymerase chain reaction (PCR) primers for regions of the genome that (i) were difficult to PCR amplify in previous studies or (ii) were found to share substantial identity with other human or mouse sequences. The combination of these new, redesigned, and existing primers sequences resulted in a total of 208,311 primer pairs (table S1)

[1]Ludwig Center for Cancer Genetics and Therapeutics, and Howard Hughes Medical Institute at Johns Hopkins Kimmel Cancer Center, Baltimore, MD 21231, USA. [2]Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston TX 77030, USA. [3]Department of Biomedical Engineering, Institute of Computational Medicine, Johns Hopkins Medical Institutions, Baltimore, MD 21218, USA. [4]Department of Neurosurgery, Johns Hopkins Medical Institutions, Baltimore, MD 21231, USA. [5]Department of Pathology, Pediatric Brain Tumor Foundation, and Preston Robert Tisch Brain Tumor Center at Duke University Medical Center, Durham, NC 27710, USA. [6]Vavilov Institute for General Genetics, Moscow B333, 117809, Russia. [7]GeneGo, Inc., St. Joseph, MI 49085, USA. [8]J. Craig Venter Institute, Rockville, MD 20850, USA. [9]Agencourt Bioscience Corporation, Beverly, MA 01915, USA. [10]Department of Neurology, School of Medicine, University of São Paulo, São Paulo, Brazil.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: bertvog@gmail.com (B.V.); velculescu@jhmi.edu (V.E.V.); kinzlke@jhmi.edu (K.W.K.)

Fig. 1. Structure of the active site of IDH1. The crystal structure of the human cytosolic NADP(+)–dependent IDH is shown in ribbon format (PDBID: 1T0L) (*44*). The active cleft of IDH1 consists of a NADP-binding site and the isocitrate-metal ion-binding site. The alpha-carboxylate oxygen and the hydroxyl group of isocitrate chelate the Ca^{2+} ion. NADP is colored in orange, isocitrate in purple and Ca^{2+} in blue. The Arg^{132} residue, displayed in yellow, forms hydrophilic interactions, shown in red, with the alpha-carboxylate of isocitrate. Displayed image was created with UCSF Chimera software version 1.2422 (*50*).
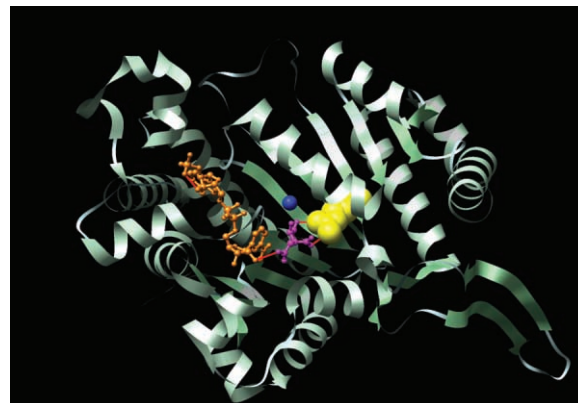
that were successfully used for sequence analysis of the coding exons of these genes.

Twenty-two GBM samples (table S2) were selected for PCR sequence analysis, consisting of 7 samples extracted directly from patient tumors and 15 samples passaged in nude mice as xenografts. In the first stage of this analysis, called the Discovery Screen, the primer pairs were used to amplify and sequence 175,471 coding exons and adjacent intronic splice donor and acceptor sequences in the 22 GBM samples and in one matched normal sample. The data were assembled for each amplified region and evaluated using stringent quality criteria (12), resulting in successful amplification and sequencing of 95.0% of targeted amplicons in the 22 tumors (Table 1). A total of 689 Mb of sequence data was generated in this fashion. The amplicon traces were analyzed using automated approaches to identify changes in the tumor sequences that were not present in the reference sequences of each gene. Alterations present in the normal control sample and in single nucleotide polymorphism (SNP) databases were then removed from further analyses. The remaining sequence traces of potential alterations were visually inspected to remove false-positive mutation calls generated by the automated software. All exons containing putative mutations were then reamplified and sequenced in both the affected tumor and the matched normal DNA sample. This process allowed us to confirm the presence of the mutation in the tumor sample and determine whether the alteration was somatic (i.e., tumor-specific) or was present in the germline. All putative somatic mutations were examined computationally and experimentally to confirm that the alterations did not arise through the aberrant coamplification of related gene sequences (12).

Analysis of sequence alterations. Analysis of the identified somatic mutations revealed that one tumor (Br27P), from a patient previously treated with radiation therapy and temozolomide,
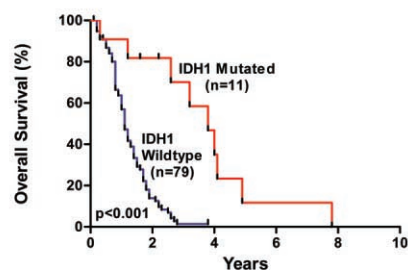
had 17 times as many alterations as any of the other 21 patients (table S3). The mutation spectrum of this sample was also dramatically different from those of the other GBM patients (12) and was consistent with previous observations of a hypermutation phenotype in glioma samples of patients treated with temozolomide (13, 14). After removing Br27P from consideration, we found that 685 genes (3.3% of the 20,661 genes analyzed) contained at least one nonsilent somatic mutation. The vast majority of these alterations were single-base substitutions (94%), whereas the others were small insertions, deletions, or duplications (Table 1). The 993 somatic mutations were observed to be distributed relatively evenly among the 21 remaining tumors (table S3), with a mean of 47 mutations per tumor, representing 1.51 mutations per Mb of GBM tumor genome sequenced. The six DNA samples extracted directly from patient tumors had smaller numbers of mutations than those obtained from xenografts, likely because of the masking effect of nonneoplastic cells in the former. It has previously been shown that cell lines and xenografts provide the optimal template DNA for cancer genome sequencing analyses (15) and that they faithfully represent the alterations present in the original tumors (16). Both the total number and the frequency of sequence alterations in GBMs were substantially smaller than the number and frequency of such alterations observed in colorectal or breast cancers and slightly less than in pancreatic cancers (10, 11, 17). The most likely explanation for this difference is the reduced

number of cell generations in glial cells before the onset of neoplasia (18).

We further evaluated a set of 21 mutated genes identified in the Discovery Screen in a second screen, called a Prevalence Screen, comprising an additional 83 GBMs with well-documented clinical histories (table S2). The 21 genes selected were mutated in at least two Discovery Screen tumors and had mutation frequencies of >10 mutations per Mb of tumor DNA sequenced. Nonsilent somatic mutations were identified in 16 of these 21 genes in the additional tumor samples (table S4). The mutation frequency of all analyzed genes in the Prevalence Screen was 23 mutations per Mb of tumor DNA, markedly increased from the overall mutation frequency in the Discovery Screen of 1.5 mutations per Mb ($P < 0.001$). Additionally, the observed ratio of nonsilent to silent mutations among mutations in the Prevalence Screen was 14.5:1, substantially higher than the 3.1:1 ratio that was observed in the Discovery Screen ($P < 0.001$). The increased mutation frequency and higher fraction of nonsilent mutations suggested that genes mutated in the Prevalence Screen were enriched for genes that actively contributed to tumorigenesis.

In addition to the frequency of mutations in a gene, the type of mutation can provide information useful for evaluating its potential role in disease (19). The likely effect of missense mutations can be assessed through evaluation of the mutated residue by evolutionary or structural means. To evaluate missense mutations, we de-

Fig. 2. Overall survival according to *IDH1* mutation status. The hazard ratio for death among patients with wild-type *IDH1* ($n = 79$), as compared to those with mutant *IDH1* ($n = 11$), was 3.7 (95 percent confidence interval, 2.1 to 6.5; $P < 0.001$). The median survival was 3.8 years for patients with mutated *IDH1*, as compared to 1.1 years for patients with wild-type *IDH1*.

Table 1. Summary of genomic analyses.

Sequencing analysis	
Number of amplicons attempted	219,229 (100%)
Number of amplicons passing quality control*	208,311 (95%)
Fraction of bases in passing amplicons with PHRED > 20	98.3%
Number of genes analyzed	20,661
Number of transcripts analyzed	23,219
Number of exons analyzed	175,471
Total number of nucleotides successfully sequenced	689,071,123
Number of somatic mutations identified ($n = 22$ samples)	2,325
Number of somatic mutations (excluding Br27P)	993
Missense	622
Nonsense	43
Insertion	3
Deletion	46
Duplication	7
Splice site or UTR	27
Synonymous	245
Average number of sequence alterations per sample	47.3
Copy number analysis	
Total number of SNP loci assessed for copy number changes	1,069,688
Number of copy number alterations identified ($n = 22$ samples)	281
Amplifications	147
Homozygous deletions	134
Average number of amplifications per sample	6.7
Average number of homozygous deletions per sample	6.1

*Passing amplicons were defined as having PHRED20 scores or better over 90% of the target sequence in 75% of samples analyzed [see (12) for additional information].

veloped an algorithm (LS-MUT) that employs machine learning of 58 predictive features based on evolutionary conservation and the physical-chemical properties of amino acids involved in the alteration (*12*). About 15% of the missense mutations evaluated were predicted to have a statistically significant effect on protein function when assessed by this method (table S3). We also were able to make structural models of 244 of the 870 missense mutations identified in this study (*20*). In each case, the model was based on x-ray crystallography or nuclear magnetic resonance spectroscopy of the normal protein or a closely related homolog. This analysis showed that 35 of the missense mutations are located close to a domain interface or substrate-binding site and thus are likely to affect protein function [links to structural models are available in (*12*)].

Analysis of copy number changes. The same tumors were then evaluated for copy number alterations through genomic hybridization of DNA samples to Illumina SNP arrays containing ~1 million probes (*21*). We have recently developed a sensitive and specific approach for the identification of focal amplifications resulting in 12 or more copies per nucleus (amplification by a factor of 6 or more compared with the diploid genome) as well as deletions of both copies of a gene (homozygous deletions) using such arrays (*22*). Unlike larger chromosomal aberrations, such focused alterations can be used to identify underlying candidate genes in these regions.

We identified a total of 147 amplifications (table S5) and 134 homozygous deletions (table S6) in the 22 samples used in the Discovery Screen (Table 1). Although the number of amplifications was similar in samples extracted from patient tumors and those that had been passaged as xenografts, the latter samples allowed detection of a larger number of homozygous deletions (average of 8.0 deletions per sample in the xenografts versus 2.2 per sample in the tumors). These observations are consistent with previous

reports that document the difficulty of identifying homozygous deletions in samples containing contaminating normal DNA (*23*) and highlight the importance of using purified human tumor cells, such as those present in xenografts or cell lines, for genomic analyses.

Integration of sequencing, copy number, and expression analyses. Mutations that arise during tumorigenesis may provide a selective advantage to the tumor cell (driver mutations) or have no net effect on tumor growth (passenger mutations). The mutational data obtained from sequencing and analysis of copy number alterations were integrated to identify GBM candidate cancer genes (*CAN*-genes) that are most likely to be drivers and therefore worthy of further investigation. To determine whether a gene was likely to harbor driver mutations, we compared the number and type of mutations observed (including sequence changes, amplifications, and homozygous deletions) and determined the probability that these alterations would result from passenger mutation rates alone (*12*) (fig. S1).

The *CAN*-genes, together with their passenger probabilities, are listed in table S7. The *CAN*-genes included several with established roles in gliomas, including *TP53*, *PTEN*, *CDKN2A*, *RB1*, *EGFR*, *NF1*, *PIK3CA*, and *PIK3R1* (*24–34*). Of these genes, the most frequently altered were *CDKN2A* (altered in 50% of GBMs); *TP53*, *EGFR*, and *PTEN* (altered in 30 to 40%); *NF1*, *CDK4*, and *RB1* (altered in 12 to 15%); and *PIK3CA* and *PIK3R1* (altered in 8 to 10%) (Table 2). Overall, these frequencies, which are similar to or in some cases higher than those previously reported, validate the sensitivity of our approach for detecting somatic alterations.

Through analysis of additional gene members within cell signaling pathways affected by these genes, we identified alterations of critical genes in the *TP53* pathway (*TP53*, *MDM2*, and *MDM4*), the *RB1* pathway (*RB1*, *CDK4*, and *CDKN2A*), and the *PI3K/PTEN* pathway

(*PIK3CA*, *PIK3R1*, *PTEN*, and *IRS1*). These alterations affected pathways in a majority of tumors (64%, 68%, and 50%, respectively), and in all cases but one, mutations within each tumor affected only a single member of each pathway in a mutually exclusive manner ($P < 0.05$) (Table 3).

Systematic analyses of functional gene groups and pathways contained within the well-annotated MetaCore database (*35*) identified enrichment of alterations in a variety of cellular processes in GBMs, including additional members of the *TP53* and *PI3K/PTEN* pathways. Many of the pathways identified were similar to core signaling pathways found to be altered in pancreas, colorectal, and breast tumors, such as those regulating control of cellular growth, apoptosis, and cell adhesion (*17*, *22*, *36*). However, several pathways were enriched only in GBMs. These included channels involved in transport of sodium, potassium, and calcium ions, as well as nervous system–specific cellular pathways such as synaptic transmission, transmission of nerve impulses, and axonal guidance (table S8). Mutations in these latter pathways may represent a subversion of normal glial cell processes to promote dysregulated growth and invasion.

Gene expression patterns can inform the analysis of pathways because they can reflect epigenetic alterations not detectable by sequencing or copy number analyses. To analyze the transcriptome of GBMs, we performed SAGE (*37*, *38*) on all GBM samples for which sufficient RNA was available (total of 16 samples), as well as on two independent normal brain RNA controls (table S9). When combined with sequencing-by-synthesis methods (*39–42*), SAGE provides a highly quantitative and sensitive measure of gene expression. We first used the transcript analysis to help identify previously uncharacterized target genes from the amplified and deleted regions that were revealed by our study. In tables S5 and S6, a candidate target gene could be identified within several of these regions

Table 2. Most frequently altered GBM *CAN*-genes. All *CAN*-genes are listed in table S7.

Gene	Point mutations*		Amplifications†		Homozygous deletions†		Fraction of tumors with any alteration (%)	Passenger probability‡
	No. of tumors	Fraction of tumors (%)	No. of tumors	Fraction of tumors (%)	No. of tumors	Fraction of tumors (%)		
CDKN2A	0/22	0	0/22	0	11/22	50	50	<0.01
TP53	37/105	35	0/22	0	1/22	5	40	<0.01
EGFR	15/105	14	5/22	23	0/22	0	37	<0.01
PTEN	27/105	26	0/22	0	1/22	5	30	<0.01
NF1	16/105	15	0/22	0	0/22	0	15	0.04
CDK4	0/22	0	3/22	14	0/22	0	14	<0.01
RB1	8/105	8	0/22	0	1/22	5	12	0.02
IDH1	12/105	11	0/22	0	0/22	0	11	<0.01
PIK3CA	10/105	10	0/22	0	0/22	0	10	0.10
PIK3R1	8/105	8	0/22	0	0/22	0	8	0.10

*Fraction of tumors with point mutations indicates the fraction of mutated GBMs out of the 105 samples in the Discovery and Prevalence Screens. CDKN2A and CDK4 were not analyzed for point mutations in the Prevalence Screen because no sequence alterations were detected in these genes in the Discovery Screen. †Fraction of tumors with amplifications and deletions indicates the number of tumors with these types of alterations in the 22 Discovery Screen samples. ‡Passenger probability indicates the probability obtained using the average of the lower and upper bound background mutation rates (*12*).

through the use of the mutational as well as transcriptional data. Second, we used the transcript analysis to help identify genes that were differentially expressed in GBMs compared to normal brain. A large number of genes (143) were expressed on average at levels 10 times as high in the GBMs. Among the overexpressed genes, 16 encoded proteins that are predicted to be secreted or expressed on the cell surface, suggesting new opportunities for diagnostic and therapeutic applications. Third, we used expression data to help assess the significance of genes containing missense mutations (table S3). Finally, we assessed whether the gene sets implicated in the pathways enriched for genetic alterations were also altered through expression changes. Notably, the gene sets in these pathways were more highly enriched for differentially expressed genes than the remaining sets ($P < 0.001$) (12). These expression data thus independently highlight the potential importance of these pathways in the development of GBMs.

High-frequency alterations of IDH1 in GBM. The *CAN*-gene list (table S7) included a number of individual genes that had not previously been linked to GBMs. The most frequently mutated of these genes, *IDH1* on chromosome 2q33, encodes isocitrate dehydrogenase 1, which catalyzes the oxidative carboxylation of isocitrate to α-ketoglutarate, resulting in the production of nicotinamide adenine dinucleotide phosphate (NADPH). Of the five isocitrate dehydrogenase proteins encoded in the human genome, at least three are localized to the mitochondria, while IDH1 is localized within the cytoplasm and peroxisomes (43). The IDH1 protein forms an asymmetric homodimer (44) and is thought to play a substantial role in cellular control of oxidative damage through generation of NADPH (45, 46). None of the other IDH genes were found to be genetically altered in our analysis.

IDH1 was somatically mutated in 5 of the 22 GBM tumors in the Discovery Screen. Surprisingly, all 5 had the same heterozygous point mutation, a change of a guanine to an adenine at position 395 of the *IDH1* transcript (G395A), leading to the replacement of an arginine with a histidine at amino acid residue 132 of the protein (R132H). In our previous study of colorectal cancers, this same codon was mutated in a single case through alteration of the adjacent nucleotide, resulting in a R132C amino acid change (10). Five GBMs evaluated in our Prevalence Screen were found to have heterozygous somatic R132H mutations, and an additional two tumors had a third distinct somatic mutation affecting the same amino acid residue, R132S (fig. S2 and Table 4). In addition to the Discovery and Prevalence Screen samples, 44 other GBMs were analyzed for *IDH1* mutations, revealing six tumors with somatic mutations affecting R132. In total, 18 of 149 GBMs (12%) analyzed had alterations in *IDH1*. The R132 residue is conserved in all known species and is localized to the substrate binding site, where it forms hydrophilic interactions with the alpha-carboxylate of isocitrate (Fig. 1) (44, 47).

Several important observations were made about *IDH1* mutations and their potential clinical importance. First, mutations in *IDH1* preferentially occurred in younger GBM patients, with a mean age of 33 years for *IDH1*-mutated patients, as opposed to 53 years for patients with wild-type *IDH1* ($P < 0.001$, t test) (Table 4). In patients under 35 years of age, nearly 50% (9 of 19) had mutations in *IDH1*. Second, mutations in *IDH1* were found in nearly all of the patients with secondary GBMs (mutations in 5 of 6 secondary GBM patients, as compared to 7 of 99 patients with primary GBMs) ($P < 0.001$, binomial test). Third, patients with *IDH1* mutations had a significantly improved prognosis, with a median overall survival of 3.8 years as compared to 1.1 years for patients with wild-type *IDH1* (Fig. 2) ($P < 0.001$, log-rank test). Although both younger age and mutated *TP53* are known to be positive prognostic factors for GBM patients, this associ-

Table 3. Mutations of the TP53, PI3K, and RB1 pathways in GBM samples. Mut, mutated; Amp, amplified; Del, deleted; Alt, altered.

Tumor sample	TP53 pathway				PI3K pathway					RB1 pathway			
	TP53	MDM2	MDM4	All genes	PTEN	PIK3CA	PIK3R1	IRS1	All genes	RB1	CDK4	CDKN2A	All genes
Br02X	Del			Alt				Mut	Alt			Del	Alt
Br03X	Mut			Alt	Mut				Alt				
Br04X	Mut			Alt	Mut				Alt	Mut			Alt
Br05X			Amp	Alt		Mut			Alt			Del	Alt
Br06X												Del	Alt
Br07X	Mut			Alt	Mut				Alt	Del			Alt
Br08X												Del	Alt
Br09P	Mut			Alt							Amp		Alt
Br10P	Mut			Alt									
Br11P	Mut			Alt									
Br12P	Mut			Alt			Mut		Alt				
Br13X	Mut			Alt								Del	Alt
Br14X							Mut		Alt			Del	Alt
Br15X										Mut		Del	Alt
Br16X		Amp		Alt							Amp		Alt
Br17X					Mut				Alt			Del	Alt
Br20P													
Br23X	Mut			Alt	Del				Alt				
Br25X					Mut				Alt			Del	Alt
Br26X						Mut			Alt			Del	Alt
Br27P	Mut			Alt							Amp		Alt
Br29P	Mut			Alt									
Fraction of tumors with altered gene/pathway*	0.55	0.05	0.05	0.64	0.27	0.09	0.09	0.05	0.50	0.14	0.14	0.45	0.68

*Fraction of affected tumors in 22 Discovery Screen samples.

ation between *IDH1* mutation and improved survival was noted even in the subgroup of young patients with *TP53* mutations (*P* < 0.02, log-rank test).

Discussion. The data resulting from this integrated analysis of mutations and copy number alterations have provided a novel view of the genetic landscape of glioblastomas. Like all large-scale genetic analyses, our study has limitations. We did not assess certain molecular alterations, including chromosomal translocations and epigenetic changes. However, our large-scale expression studies should have identified any genes that were differentially expressed through these mechanisms (table S9). Additionally, we focused on copy number changes that were focal amplifications or homozygous deletions, because these have historically been most useful in identifying cancer genes. The array data we have generated can also be analyzed to determine loss of heterozygosity (LOH) or low-amplitude regions of copy number gains, but such changes cannot generally be used to pinpoint new candidate cancer genes. Finally, the samples directly extracted from patient tumors contained small amounts of contaminating normal tissue, which limited our ability to detect homozygous deletions and, to a lesser extent, somatic mutations, in those specific tumors.

Despite these limitations, our study provides a number of important genetic and clinical insights into GBMs. First, it revealed that some of the pathways known to be altered in GBMs affect a larger fraction of genes and patients than previously anticipated. A majority of the tumors analyzed had alterations in genes encoding components of each of the *TP53*, *RB1*, and *PI3K* pathways. The fact that all but one of the cancers with mutations in members of a pathway did not have alterations in other members of the same pathway suggests that such alterations are functionally equivalent in tumorigenesis. Second, these results have identified a variety of new genes and signaling pathways not previously implicated in GBMs (table S7 and S8). Some of these pathways were found to be altered in previous genome-wide analyses of pancreatic, breast, and colorectal cancers and may represent core processes that underlie human tumorigenesis (*17, 22, 36*). A number of the signaling pathways mutated or altered through expression differences in GBMs appear to be involved in nervous system signaling processes and represent novel and potentially useful aspects of GBM biology.

The comprehensive nature of our study allowed us to identify *IDH1* as an unexpected target of genetic alteration in patients with GBM. All mutations in this gene resulted in amino acid substitutions at position 132, an evolutionarily conserved residue located within the isocitrate binding site (*44*). The recurrent nature of the mutations is reminiscent of activating alterations in oncogenes such as *BRAF*, *KRAS*, and *PIK3CA*. Our speculation that this sequence change is an activating mutation is strengthened by the absence of inactivating changes (e.g., frameshift or stop mutations), the absence of other alterations in key residues of the active site, and the fact that all mutations observed to date were heterozygous (without any evidence of loss of the second allele through LOH). Interestingly, enzymatic studies have shown that in vitro engineered substitution of arginine at residue 132 with a different amino acid (glutamate) than that observed in patients results in a catalytically inactive enzyme, suggesting a critical role for this residue (*48*). Further biochemical and molecular analyses will be needed to determine the effect of alterations of IDH1 on enzymatic activity and cellular phenotype.

Regardless of the specific molecular consequences of *IDH1* alterations, detection of mutations in *IDH1* is likely to be clinically useful. Although considerable effort has focused on the identification of characteristic genetic lesions in primary and secondary GBMs, the altered genes identified to date are not optimal for this purpose (*5*). Our study revealed IDH1 mutation to be a novel and potentially more specific marker for secondary GBM. One hypothesis is that *IDH1* alterations identify a biologically specific subgroup of GBM patients, including both patients who would be classified as having secondary GBMs and a subpopulation of primary GBM patients with a similar tumor biology and a more protracted clinical course (Table 4). Interestingly, patients with *IDH1* mutations had a very high frequency of *TP53* mutation and a very low frequency of mutations in other commonly altered GBM genes (Table 4). Patients with mutated

Table 4. Characteristics of GBM patients with IDH1 mutations

Patient ID	Patient age (years)*	Sex	Recurrent GBM†	Secondary GBM‡	Overall survival (years)§	IDH1 mutation		Mutation of TP53	Mutation of PTEN, RB1, EGFR, or NF1
						Nucleotide	Amino acid		
Br10P	30	F	No	No	2.2	G395A	R132H	Yes	No
Br11P	32	M	No	No	4.1	G395A	R132H	Yes	No
Br12P	31	M	No	No	1.6	G395A	R132H	Yes	No
Br104X	29	F	No	No	4.0	C394A	R132S	Yes	No
Br106X	36	M	No	No	3.8	G395A	R132H	Yes	No
Br122X	53	M	No	No	7.8	G395A	R132H	No	No
Br123X	34	M	No	Yes	4.9	G395A	R132H	Yes	No
Br237T	26	M	No	Yes	2.6	G395A	R132H	Yes	No
Br211T	28	F	No	Yes	0.3	G395A	R132H	Yes	No
Br27P	32	M	Yes	Yes	1.2	G395A	R132H	Yes	No
Br129X	25	M	Yes	Yes	3.2	C394A	R132S	No	No
Br29P	42	F	Yes	Unknown	Unknown	G395A	R132H	Yes	No
IDH1 mutant patients (*n*=12)	33.2	67% M	25%	42%	3.8	100%	100%	83%	0%
IDH1 wild-type patients (*n*=93)	53.3	65% M	16%	1%	1.1	0%	0%	27%	60%

*Patient age refers to age at which the sample was obtained.   †Recurrent GBM designates a GBM which was resected >3 months after a prior diagnosis of GBM.   ‡Secondary GBM designates a GBM which was resected > 1 year after a prior diagnosis of a lower grade glioma (WHO I-III).   §Overall survival was calculated using date of GBM diagnosis and date of death or last patient contact: Patients Br10P and Br11P were alive at last contact. Median survival for IDH1 mutant patients and IDH1 wild-type patients was calculated using logrank test. Previous pathologic diagnoses in secondary GBM patients were oligodendroglioma (WHO grade II) in Br123X, low grade glioma (WHO grade I-II) in Br237T and Br211T, anaplastic astrocytoma (WHO grade III) in Br27P, and anaplastic oligodendroglioma (WHO grade III) in Br129X. Mean age and median survival are listed for the groups of IDH1-mutated and IDH1-wild-type patients.

IDH1 also had distinct clinical characteristics, including younger age and a considerably improved clinical prognosis (Table 4). It is conceivable that new treatments could be designed to take advantage of IDH1 alterations in these patients, because inhibition of a different IDH enzyme (IDH2) has recently been shown to result in increased sensitivity of tumor cells to a variety of chemotherapeutic agents (49). In summary, the discovery of IDH1 and other genes previously not known to play a role in human tumors (table S7) validates the utility of genome-wide genetic analysis of tumors in general and opens new avenues of basic and clinical brain tumor research.

References and Notes

1. D. N. Louis et al., Acta Neuropathol. **114**, 97 (2007).
2. R. Stupp et al., N. Engl. J. Med. **352**, 987 (2005).
3. H. Scherer, Am. J. Cancer **40**, 159 (1940).
4. P. Kleihues, H. Ohgaki, Neuro-oncol. **1**, 44 (1999).
5. H. Ohgaki, P. Kleihues, Am. J. Pathol. **170**, 1445 (2007).
6. H. Ohgaki et al., Cancer Res. **64**, 6892 (2004).
7. I. K. Mellinghoff et al., N. Engl. J. Med. **353**, 2012 (2005).
8. E. A. Maher et al., Cancer Res. **66**, 11502 (2006).
9. C. L. Tso et al., Cancer Res. **66**, 159 (2006).
10. T. Sjöblom et al., Science **314**, 268 (2006).
11. L. D. Wood et al., Science **318**, 1108 (2007).
12. Materials and methods are available as supporting material on Science Online.
13. D. P. Cahill et al., Clin. Cancer Res. **13**, 2038 (2007).
14. C. Hunter et al., Cancer Res. **66**, 3987 (2006).
15. J. M. Winter, J. R. Brody, S. E. Kern, Cancer Biol. Ther. **5**, 360 (2006).
16. S. Jones et al., Proc. Natl. Acad. Sci. U.S.A. **105**, 4283 (2008).
17. S. Jones et al., Science **321**, 1801 (2008).
18. R. Kraus-Ruppert, J. Laissue, H. Burki, N. Odartchenko, J. Comp. Neurol. **148**, 211 (1973).
19. P. C. Ng, S. Henikoff, Nucleic Acids Res. **31**, 3812 (2003).
20. R. Karchin, Structural models of mutants identified in glioblastomas (2008); http://karchinlab.org/Mutants/CAN-genes/brain/GBM.html
21. F. J. Steemers et al., Nat. Methods **3**, 31 (2006).
22. R. J. Leary et al., Proc. Natl. Acad. Sci. U.S.A., in press.
23. P. Cairns et al., Nat. Genet. **11**, 210 (1995).
24. J. M. Nigro et al., Nature **342**, 705 (1989).
25. J. Li et al., Science **275**, 1943 (1997).
26. K. Ueki et al., Cancer Res. **56**, 150 (1996).
27. A. J. Wong et al., Proc. Natl. Acad. Sci. U.S.A. **84**, 6899 (1987).
28. M. Mizoguchi, C. L. Nutt, G. Mohapatra, D. N. Louis, Brain Pathol. **14**, 372 (2004).
29. L. Frederick, X. Y. Wang, G. Eley, C. D. James, Cancer Res. **60**, 1383 (2000).
30. Y. Li et al., Cell **69**, 275 (1992).
31. G. Thiel et al., Anticancer Res. **15**, 2495 (1995).
32. Y. Samuels et al., Science **304**, 554 (2004).
33. D. K. Broderick et al., Cancer Res. **64**, 5048 (2004).
34. G. L. Gallia et al., Mol. Cancer Res. **4**, 709 (2006).
35. S. Ekins, Y. Nikolsky, A. Bugrim, E. Kirillov, T. Nikolskaya, Methods Mol. Biol. **356**, 319 (2007).
36. J. Lin et al., Genome Res. **17**, 1304 (2007).
37. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, Science **270**, 484 (1995).
38. S. Saha et al., Nat. Biotechnol. **20**, 508 (2002).
39. M. Sultan et al., Science **321**, 956 (2008).
40. R. Lister et al., Cell **133**, 523 (2008).
41. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, Nat. Methods **5**, 621 (2008).
42. R. Morin et al., Biotechniques **45**, 81 (2008).
43. B. V. Geisbrecht, S. J. Gould, J. Biol. Chem. **274**, 30527 (1999).
44. X. Xu et al., J. Biol. Chem. **279**, 33946 (2004).
45. S. M. Lee et al., Free Radic. Biol. Med. **32**, 1185 (2002).
46. S. Y. Kim et al., Mol. Cell. Biochem. **302**, 27 (2007).
47. A. Nekrutenko, D. M. Hillis, J. C. Patton, R. D. Bradley, R. J. Baker, Mol. Biol. Evol. **15**, 1674 (1998).
48. G. T. Jennings, K. I. Minard, L. McAlister-Henn, Biochemistry **36**, 13743 (1997).
49. I. S. Kil, S. Y. Kim, S. J. Lee, J. W. Park, Free Radic. Biol. Med. **43**, 1197 (2007).
50. E. F. Pettersen et al., J. Comput. Chem. **25**, 1605 (2004).
51. Contact information for the authors who directed the major components of this project is as follows: H. Yan, G. J. Riggins (clinical and sample coordination), yan00002@mc.duke.edu; griggin1@jhmi.edu; D. D. Bigner (pathological and clinical review), bigne001@mc.duke.

REPORTS

Quantum Communication with Zero-Capacity Channels

Graeme Smith[1]* and Jon Yard[2]

Communication over a noisy quantum channel introduces errors in the transmission that must be corrected. A fundamental bound on quantum error correction is the quantum capacity, which quantifies the amount of quantum data that can be protected. We show theoretically that two quantum channels, each with a transmission capacity of zero, can have a nonzero capacity when used together. This unveils a rich structure in the theory of quantum communications, implying that the quantum capacity does not completely specify a channel's ability to transmit quantum information.

Noise is the enemy of all modern communication links. Cellular, Internet and satellite communications all depend crucially on active steps taken to mitigate and correct for noise. The study of communication in the presence of noise was formalized by Shannon (1), who simplified the analysis by making probabilistic assumptions about the nature of the noise. By modeling a noisy channel $\mathcal{N}$ as a probabilistic map from input signals to output signals, the capacity $\mathcal{C}(\mathcal{N})$ of $\mathcal{N}$ is defined as the number of bits that can be transmitted per channel use, with vanishing errors in the limit of many transmissions. This capacity is computed via the formula $\mathcal{C}(\mathcal{N}) = \max_X I(X; Y)$, where the maximization is over random variables X at the input of the channel, Y is the resulting output of the channel, and the mutual information $I(X;Y) = H(X) + H(Y) - H(X,Y)$ quantifies the correlation between input and output. $H(X) = -\sum_x p_x \log_2 p_x$ denotes the Shannon entropy, which quantifies the amount of randomness in X. The capacity, measured in bits per channel use, is the fundamental bound between communication rates that are achievable in principle and those that are not. The capacity formula guides the design of practical error-correction techniques by providing a benchmark against which engineers can test the performance of their systems. Practical implementations guided by the capacity result now come strikingly close to the Shannon limit (2).

A fundamental prediction of the capacity formula is that the only channels with zero capacity are precisely those for which the input

Quantum Communication with Zero-Capacity Channels

Graeme Smith, *et al.*
Science **321**, 1812 (2008);
DOI: 10.1126/science.1162242

IDH1 also had distinct clinical characteristics, including younger age and a considerably improved clinical prognosis (Table 4). It is conceivable that new treatments could be designed to take advantage of IDH1 alterations in these patients, because inhibition of a different IDH enzyme (IDH2) has recently been shown to result in increased sensitivity of tumor cells to a variety of chemotherapeutic agents (49). In summary, the discovery of IDH1 and other genes previously not known to play a role in human tumors (table S7) validates the utility of genome-wide genetic analysis of tumors in general and opens new avenues of basic and clinical brain tumor research.

References and Notes

1. D. N. Louis *et al.*, *Acta Neuropathol.* **114**, 97 (2007).
2. R. Stupp *et al.*, *N. Engl. J. Med.* **352**, 987 (2005).
3. H. Scherer, *Am. J. Cancer* **40**, 159 (1940).
4. P. Kleihues, H. Ohgaki, *Neuro-oncol.* **1**, 44 (1999).
5. H. Ohgaki, P. Kleihues, *Am. J. Pathol.* **170**, 1445 (2007).
6. H. Ohgaki *et al.*, *Cancer Res.* **64**, 6892 (2004).
7. I. K. Mellinghoff *et al.*, *N. Engl. J. Med.* **353**, 2012 (2005).
8. E. A. Maher *et al.*, *Cancer Res.* **66**, 11502 (2006).
9. C. L. Tso *et al.*, *Cancer Res.* **66**, 159 (2006).
10. T. Sjöblom *et al.*, *Science* **314**, 268 (2006).
11. L. D. Wood *et al.*, *Science* **318**, 1108 (2007).
12. Materials and methods are available as supporting material on *Science* Online.
13. D. P. Cahill *et al.*, *Clin. Cancer Res.* **13**, 2038 (2007).
14. C. Hunter *et al.*, *Cancer Res.* **66**, 3987 (2006).
15. J. M. Winter, J. R. Brody, S. E. Kern, *Cancer Biol. Ther.* **5**, 360 (2006).
16. S. Jones *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4283 (2008).
17. S. Jones *et al.*, *Science* **321**, 1801 (2008).
18. R. Kraus-Ruppert, J. Laissue, H. Burki, N. Odartchenko, *J. Comp. Neurol.* **148**, 211 (1973).
19. P. C. Ng, S. Henikoff, *Nucleic Acids Res.* **31**, 3812 (2003).
20. R. Karchin, Structural models of mutants identified in glioblastomas (2008); http://karchinlab.org/Mutants/CAN-genes/brain/GBM.html

21. F. J. Steemers *et al.*, *Nat. Methods* **3**, 31 (2006).
22. R. J. Leary *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
23. P. Cairns *et al.*, *Nat. Genet.* **11**, 210 (1995).
24. J. M. Nigro *et al.*, *Nature* **342**, 705 (1989).
25. J. Li *et al.*, *Science* **275**, 1943 (1997).
26. K. Ueki *et al.*, *Cancer Res.* **56**, 150 (1996).
27. A. J. Wong *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6899 (1987).
28. M. Mizoguchi, C. L. Nutt, G. Mohapatra, D. N. Louis, *Brain Pathol.* **14**, 372 (2004).
29. L. Frederick, X. Y. Wang, G. Eley, C. D. James, *Cancer Res.* **60**, 1383 (2000).
30. Y. Li *et al.*, *Cell* **69**, 275 (1992).
31. G. Thiel *et al.*, *Anticancer Res.* **15**, 2495 (1995).
32. Y. Samuels *et al.*, *Science* **304**, 554 (2004).
33. D. K. Broderick *et al.*, *Cancer Res.* **64**, 5048 (2004).
34. G. L. Gallia *et al.*, *Mol. Cancer Res.* **4**, 709 (2006).
35. S. Ekins, Y. Nikolsky, A. Bugrim, E. Kirillov, T. Nikolskaya, *Methods Mol. Biol.* **356**, 319 (2007).
36. J. Lin *et al.*, *Genome Res.* **17**, 1304 (2007).
37. V. E. Velculescu, L. Zhang, B. Vogelstein, K. W. Kinzler, *Science* **270**, 484 (1995).
38. S. Saha *et al.*, *Nat. Biotechnol.* **20**, 508 (2002).
39. M. Sultan *et al.*, *Science* **321**, 956 (2008).
40. R. Lister *et al.*, *Cell* **133**, 523 (2008).
41. A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nat. Methods* **5**, 621 (2008).
42. R. Morin *et al.*, *Biotechniques* **45**, 81 (2008).
43. B. V. Geisbrecht, S. J. Gould, *J. Biol. Chem.* **274**, 30527 (1999).
44. X. Xu *et al.*, *J. Biol. Chem.* **279**, 33946 (2004).
45. S. M. Lee *et al.*, *Free Radic. Biol. Med.* **32**, 1185 (2002).
46. S. Y. Kim *et al.*, *Mol. Cell. Biochem.* **302**, 27 (2007).
47. A. Nekrutenko, D. M. Hillis, J. C. Patton, R. D. Bradley, R. J. Baker, *Mol. Biol. Evol.* **15**, 1674 (1998).
48. G. T. Jennings, K. I. Minard, L. McAlister-Henn, *Biochemistry* **36**, 13743 (1997).
49. I. S. Kil, S. Y. Kim, S. J. Lee, J. W. Park, *Free Radic. Biol. Med.* **43**, 1197 (2007).
50. E. F. Pettersen *et al.*, *J. Comput. Chem.* **25**, 1605 (2004).
51. Contact information for the authors who directed the major components of this project is as follows: H. Yan, G. J. Riggins (clinical and sample coordination), yan00002@mc.duke.edu, griggin1@jhmi.edu; D. D. Bigner (pathological and clinical review), bigne001@mc.duke.

REPORTS

Quantum Communication with Zero-Capacity Channels

Graeme Smith[1]* and Jon Yard[2]

Communication over a noisy quantum channel introduces errors in the transmission that must be corrected. A fundamental bound on quantum error correction is the quantum capacity, which quantifies the amount of quantum data that can be protected. We show theoretically that two quantum channels, each with a transmission capacity of zero, can have a nonzero capacity when used together. This unveils a rich structure in the theory of quantum communications, implying that the quantum capacity does not completely specify a channel's ability to transmit quantum information.

Noise is the enemy of all modern communication links. Cellular, Internet and satellite communications all depend crucially on active steps taken to mitigate and correct for noise. The study of communication in the presence of noise was formalized by Shannon (1), who simplified the analysis by making probabilistic assumptions about the nature of the noise. By modeling a noisy channel $\mathcal{N}$ as a probabilistic map from input signals to output signals, the capacity $\mathcal{C}(\mathcal{N})$ of $\mathcal{N}$ is defined as the number of bits that can be transmitted per channel use, with vanishing errors in the limit of many transmissions. This capacity is computed via the formula $\mathcal{C}(\mathcal{N}) = \max_X I(X; Y)$, where the maximization is over random variables X at the input of the channel, Y is the resulting output of the channel, and the mutual information $I(X;Y) = H(X) + H(Y) - H(X,Y)$ quantifies the correlation between input and output. $H(X) = -\sum_x p_x \log_2 p_x$ denotes the Shannon entropy, which quantifies the amount of randomness in X. The capacity, measured in bits per channel use, is the fundamental bound between communication rates that are achievable in principle and those that are not. The capacity formula guides the design of practical error-correction techniques by providing a benchmark against which engineers can test the performance of their systems. Practical implementations guided by the capacity result now come strikingly close to the Shannon limit (2).

A fundamental prediction of the capacity formula is that the only channels with zero capacity are precisely those for which the input

and output are completely uncorrelated. Furthermore, suppose one is given simultaneous access to two noisy channels $\mathcal{N}_1$ and $\mathcal{N}_2$. The capacity of the product channel $\mathcal{N}_1 \times \mathcal{N}_2$, where the channels are used in parallel, takes the simple form $\mathcal{C}(\mathcal{N}_1 \times \mathcal{N}_2) = \mathcal{C}(\mathcal{N}_1) + \mathcal{C}(\mathcal{N}_2)$; that is, the capacity is additive. Additivity shows that capacity is an intrinsic measure of the information-conveying properties of a channel.

Quantum data are an especially delicate form of information and are particularly susceptible to the deleterious effects of noise. Because quantum communication promises to allow unconditionally secure communication (3), and a quantum computer could dramatically speed up some computations (4), there is tremendous interest in techniques to protect quantum data from noise. A quantum channel $\mathcal{N}$ models a physical process that adds noise to a quantum system via an interaction with an unobservable environment (Fig. 1), generalizing Shannon's model and enabling a more accurate depiction of the underlying physics. In this setting, it is natural to ask what the capacity of a quantum channel is for transmitting quantum-mechanical information (5) and whether it has a simple formula in analogy with Shannon's.

Just as any classical message can be reversibly expressed as a sequence of bits, a quantum message (that is, an arbitrary state of a given quantum system) can be reversibly transferred to a collection of two-level quantum systems, or "qubits," giving a measure of the size of the system. The goal of quantum communication is to transfer the joint state of a collection of qubits from one location to another (Fig. 2). The quantum capacity $\mathcal{Q}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is the number of qubits per channel use that can be reliably transmitted via many noisy transmissions, where each transmission is modeled by $\mathcal{N}$. Although noiseless quantum communication with a noisy quantum channel is one of the simplest and most natural communication tasks one can imagine for quantum information, it is not nearly as well understood as its classical counterpart.

An analog for mutual information in the quantum capacity has been proposed (6) and called the "coherent information"

$$\mathcal{Q}^{(1)}(\mathcal{N}) = \max_{\rho^A}[H(B) - H(E)] \qquad (1)$$

The entropies are measured on the states induced at the output and environment of the channel (Fig. 1) by the input state ρ^A, where $H(B)$ is the von Neumann entropy of the state ρ^B at the output. Coherent information is rather different from mutual information. This difference is

closely related to the no-cloning theorem (7), which states that quantum information cannot be copied, because the coherent information roughly measures how much more information B holds than E. The no-cloning theorem itself is deeply tied to the fundamentally quantum concept of entanglement, in which the whole of a quantum system can be in a definite state while the states of its parts are uncertain.

The best-known expression for the quantum capacity $\mathcal{Q}$ is given (8–10) by the "regularization" of $\mathcal{Q}^{(1)}$: $\mathcal{Q}(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n}\mathcal{Q}^{(1)}(\mathcal{N}^{\times n})$. Here $\mathcal{N}^{\times n}$ represents the parallel use of n copies of $\mathcal{N}$. The asymptotic nature of this expression prevents one from determining the quantum capacity of a given channel in any effective way, while also making it difficult to reason about its general properties. In contrast to Shannon's capacity, where regularization is unnecessary, here it cannot be removed in general (11, 12). Consequently, even apparently simple questions, such as determining from a channel's description whether it can be used to send any quantum information, are currently unresolved. We find that the answer to this question depends on context; there are pairs of zero-capacity channels that, used together, have a positive quantum capacity (Fig. 3). This shows that the quantum capacity is not additive, and thus the quantum capacity of a channel does not completely specify its capability for transmitting quantum information.

Although a complete characterization of zero-capacity channels is unknown, certain classes of

zero-capacity channels are known. One class consists of channels for which the joint quantum state of the output and environment is symmetric under interchange. These symmetric channels are quite different from Shannon's zero-capacity channels, because they display correlations between the input and output. However, they are useless by themselves for quantum communication because their symmetry implies that any capacity would lead to a violation of the no-cloning theorem (7, 13). Another class of zero-capacity channels is entanglement-binding channels (14, 15), also called Horodecki channels, which can only produce very weakly entangled states, satisfying a condition called positive partial transposition (16).

Even though channels from one or the other of these classes cannot be combined to faithfully transmit quantum data, we find that when one combines a channel from each class, it is sometimes possible to obtain a positive quantum capacity. We do this by proving a new relationship between two further capacities of a quantum channel: the private capacity (10) and the assisted capacity (17).

The private capacity $\mathcal{P}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ is the rate at which it can be used to send classical data that is secure against an eavesdropper with access to the environment of the channel. This capacity is closely related to quantum key distribution protocols (3) and was shown (10) to equal the regularization of the private information

$$\mathcal{P}^{(1)}(\mathcal{N}) = \max_{X, \rho_x^A}(I(X;B) - I(X;E)) \qquad (2)$$
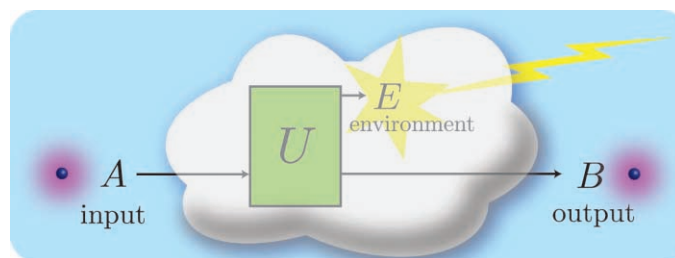
Fig. 1. Representation of a quantum channel. A channel reversibly transfers the state of a physical system in the laboratory of the sender to the combination of the system possessed by the receiver and an environment that is inaccessible to the users of the channel. Discarding the environment results in a noisy evolution of the state. The input and output denote separate places in space and/or time, modeling, for example, a leaky optical fiber or the irreversible evolution of the state of a quantum dot.
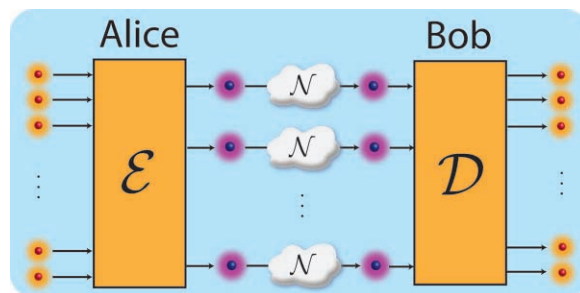
Fig. 2. The quantum capacity of a quantum channel. Quantum data are held by a sender (traditionally called Alice), who would like to transmit it to a receiver (Bob) with many parallel uses of a noisy quantum channel $\mathcal{N}$. Alice encodes the data with a collective encoding operation $\mathcal{E}$, which results in a joint quantum state on the inputs of the channels $\mathcal{N}^{\times n}$. The encoded state is sent through the noisy channels. When Bob receives the state, he applies a decoding operation $\mathcal{D}$, which acts collectively on the many outputs of the channels. After decoding, Bob holds the state that Alice wished to send. The quantum capacity is the total number of qubits in the state Alice sends divided by the number of channel uses.

[1]IBM T. J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598, USA. [2]Quantum Institute, Center for Nonlinear Studies (CNLS), Computer, Computational and Statistical Sciences (CCS-3), Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

*To whom correspondence should be addressed. E-mail: gsbsmith@gmail.com
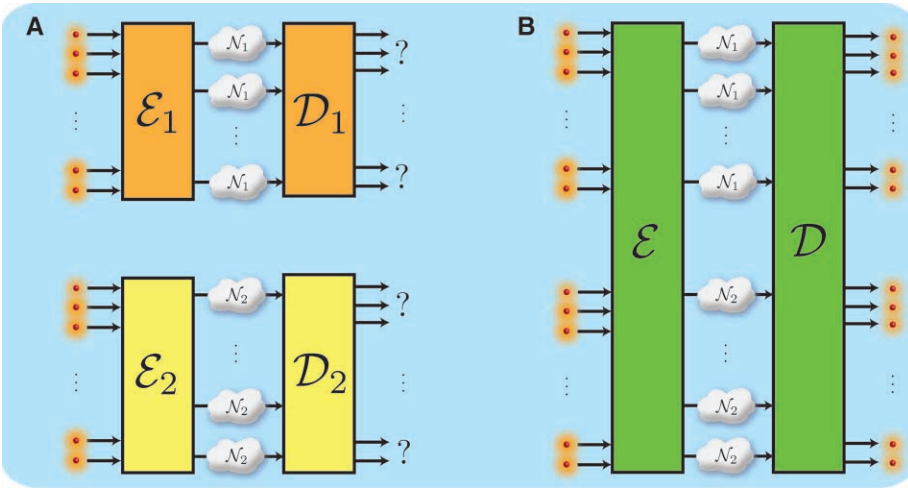
Fig. 3. (**A**) Alice and Bob attempt to separately use two zero-capacity channels $\mathcal{N}_1$ and $\mathcal{N}_2$ to transfer quantum states. Alice uses separate encoders $\mathcal{E}_1$ and $\mathcal{E}_2$ for each group of channels, and Bob uses separate decoders $\mathcal{D}_1$ and $\mathcal{D}_2$. Any attempt will fail because the capacity of each channel is zero. (**B**) The same two channels being used in parallel for the same task. Alice's encoder $\mathcal{E}$ now has simultaneous access to the inputs of all channels being used and Bob's decoding $\mathcal{D}$ is also performed jointly. Noiseless communication is nonetheless possible because $\mathcal{Q}$ is not additive.

Fig. 4. Relating the private capacity and the assisted capacity. A straightforward proof of Eq. 3 uses the expression (17) $\mathcal{Q}_{\mathcal{A}}(\mathcal{N}) = \frac{1}{2}\max_{\rho^{XAC}}[I(X;B|C) - I(X;E|C)]$. Here, $I(X;B|C)$ is the conditional mutual information: $H(XC) + H(BC) - H(XBC) - H(C)$. It is evaluated on the state obtained by putting the $\mathcal{A}$ part of a state ρ^{XAC} into the channel $\mathcal{N}$, which can be thought of as mapping $A \to BE$ as in Fig. 2. The maximization here is similar in form to Eq. 2, but is over a less-constrained type of state. Therefore, $\frac{1}{2}\mathcal{P}^{(1)}(\mathcal{N}) \leq \mathcal{Q}_{\mathcal{A}}(\mathcal{N})$. This bound holds for the associated regularized quantities, and because regularization does not change $\mathcal{Q}_{\mathcal{A}}$, Eq. 3 follows.

where the maximization is over classical random variables X and quantum states ρ_x^A on the input of $\mathcal{N}$ depending on the value x of X.

In order to find upper bounds on the quantum capacity, an "assisted capacity" was recently introduced (17), in which one allows the free use of arbitrary symmetric channels to assist quantum communication over a given channel. Letting $\mathcal{A}$ denote a symmetric channel of unbounded dimension (the strongest such channel), the assisted capacity $\mathcal{Q}_{\mathcal{A}}(\mathcal{N})$ of a quantum channel $\mathcal{N}$ satisfies (17) $\mathcal{Q}_{\mathcal{A}}(\mathcal{N}) = \mathcal{Q}(\mathcal{N} \times \mathcal{A}) = \mathcal{Q}^{(1)}(\mathcal{N} \times \mathcal{A})$.

Because the dimension of the input to $\mathcal{A}$ is unbounded, we cannot evaluate the assisted capacity in general. Nonetheless, the assisted capacity helps to reason about finite-dimensional channels.

Although Horodecki channels have zero quantum capacity, examples of such channels with nonzero private capacity are known (18, 19). One of the two zero-capacity channels we will combine to give positive joint capacity is such a private Horodecki channel $\mathcal{N}_H$ and the other is the symmetric channel $\mathcal{A}$. Our key tool is the following new relationship between the capacities of any channel $\mathcal{N}$ (Fig. 4)

$$\frac{1}{2}\mathcal{P}(\mathcal{N}) \leq \mathcal{Q}_{\mathcal{A}}(\mathcal{N}) \qquad (3)$$

A channel's assisted capacity is at least as large as half its private capacity. It follows that any private Horodecki channel $\mathcal{N}_H$ has a positive assisted capacity, and thus the two zero-capacity channels $\mathcal{N}_H$ and $\mathcal{A}$ satisfy $\mathcal{Q}_{\mathcal{A}}(\mathcal{N}_H) = \mathcal{Q}(\mathcal{N}_H \times \mathcal{A}) > 0$.

Although our construction involves systems of unbounded dimension, one can show that any private Horodecki channel can be combined with a finite symmetric channel to give positive

quantum capacity. In particular, there is a private Horodecki channel acting on a four-level system (19). This channel gives positive quantum capacity when combined with a small symmetric channel—a 50% erasure channel $\mathcal{A}_e$ with a four-level input, which half of the time delivers the input state to the output, otherwise telling the receiver that an erasure has occurred. We show (20) that the parallel combination of these channels has a quantum capacity greater than 0.01.

We find this "superactivation" to be a startling effect. One would think that the question, "can this communication link transmit any information?" would have a straightforward answer. However, with quantum data, the answer may well be "it depends on the context." Taken separately, private Horodecki channels and symmetric channels are useless for transmitting quantum information, albeit for entirely different reasons. Nonetheless, each channel has the potential to activate the other, effectively canceling the other's reason for having zero capacity. We know of no analog of this effect in the classical theory. Perhaps each channel transfers some different but complementary kind of quantum information. If so, can these kinds of information be quantified in an operationally meaningful way? Are there other pairs of zero-capacity channels displaying this effect? Are there triples? Does the private capacity also display superactivation? Can all Horodecki channels be superactivated, or just those with positive private capacity? What new insights does this yield for computing the quantum capacity in general?

Besides additivity, our findings resolve two open questions about the quantum capacity. First we find (20) that the quantum capacity is not a convex function of the channel. Convexity of a capacity means that a probabilistic mixture of

two channels never has a higher capacity than the corresponding average of the capacities of the individual channels. Violation of convexity leads to a counterintuitive situation in which it can be beneficial to forget which channel is being used. We also find (20) channels with an arbitrarily large gap between $\mathcal{Q}^{(1)}$ [the so-called "hashing rate" (8–10)] and the quantum capacity. It had been consistent with previous results (11, 12) to believe that $\mathcal{Q}$ and $\mathcal{Q}^{(1)}$ would be equal up to small corrections. Our work shows that this is not the case and indicates that the hashing rate is an overly pessimistic benchmark against which to measure the performance of practical error-correction schemes. This could be good news for the analysis of fault-tolerant quantum computation in the very noisy regime.

Forms of this sort of superactivation are known in the multiparty setting, where several separated parties communicate via a quantum channel with multiple inputs or outputs (21–24), and have been conjectured for a quantum channel assisted by classical communication between the sender and receiver (25, 26). Because these settings are rather complex, it is perhaps unsurprising to find exotic behavior. In contrast, the problem of noiseless quantum communication with a noisy quantum channel is one of the simplest and most natural communication tasks imaginable in a quantum-mechanical context. Our findings uncover a level of complexity in this simple problem that had not been anticipated and point toward several fundamentally new questions about information and communication in the physical world.

References and Notes
1. C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
2. T. Richardson, R. Urbanke, *IEEE Commun. Mag.* **41**, 126 (2003).
3. C. H. Bennett, G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, New York, 1984), pp. 175–179.

4. P. W. Shor, *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Los Alaminos, CA 1994), pp. 124–134.
5. C. H. Bennett, P. W. Shor, *Science* **303**, 1784 (2004).
6. B. Schumacher, M. A. Nielsen, *Phys. Rev. A* **54**, 2629 (1996).
7. W. Wootters, W. Zurek, *Nature* **299**, 802 (1982).
8. S. Lloyd, *Phys. Rev. A* **55**, 1613 (1997).
9. P. W. Shor, lecture notes from the Mathematical Sciences Research Institute Workshop on Quantum Computation, Berkeley, CA, 2002. Available online at www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/.
10. I. Devetak, *IEEE Trans. Inf. Theory* **51**, 44 (2005).
11. D. DiVincenzo, P. W. Shor, J. A. Smolin, *Phys. Rev. A* **57**, 830 (1998).
12. G. Smith, J. A. Smolin, *Phys. Rev. Lett.* **98**, 030501 (2007).
13. C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, *Phys. Rev. Lett.* **78**, 3217 (1997).
14. M. Horodecki, P. Horodecki, R. Horodecki, *Phys. Lett. A* **223**, 1 (1996).
15. P. Horodecki, *Phys. Lett. A* **232**, 333 (1997).
16. A. Peres, *Phys. Rev. Lett.* **77**, 1413 (1996).
17. G. Smith, J. Smolin, A. Winter, *IEEE Trans. Inf. Theory* **54**, 4208 (2008).
18. K. Horodecki, M. Horodecki, P. Horodecki, J. Oppenheim, *Phys. Rev. Lett.* **94**, 160502 (2005).
19. K. Horodecki, L. Pankowski, M. Horodecki, P. Horodecki, *IEEE Trans. Inf. Theory* **54**, 2621 (2008).
20. Further details can be found in the supporting online material on *Science* Online.
21. P. Shor, J. Smolin, A. Thapliyal, *Phys. Rev. Lett.* **90**, 107901 (2003).
22. W. Dur, J. Cirac, P. Horodecki, *Phys. Rev. Lett.* **93**, 020503 (2004).
23. R. Duan, Y. Shi, *Phys. Rev. Lett.* **101**, 020501 (2008).
24. L. Czekaj, P. Horodecki, preprint available at http://arxiv.org/pdf/0807.3977.
25. P. Horodecki, M. Horodecki, R. Horodecki, *Phys. Rev. Lett.* **82**, 1056 (1999).
26. P. Shor, J. Smolin, B. Terhal, *Phys. Rev. Lett.* **86**, 2681 (2001).
27. We are indebted to C. Bennett, C. Callaway, E. Timmermans, B. Toner, and A. Winter for encouragement and comments on an earlier draft. J.Y. is supported by the CNLS and the Quantum Institute through grants provided by the Laboratory Directed Research and Development program of the U.S. Department of Energy.

Synthesis and Solid-State NMR Structural Characterization of ^{13}C-Labeled Graphite Oxide

Weiwei Cai,[1,2] Richard D. Piner,[1] Frank J. Stadermann,[3] Sungjin Park,[1] Medhat A. Shaibat,[4] Yoshitaka Ishii,[4] Dongxing Yang,[1] Aruna Velamakanni,[1] Sung Jin An,[5] Meryl Stoller,[1] Jinho An,[1] Dongmin Chen,[2] Rodney S. Ruoff[1]*

The detailed chemical structure of graphite oxide (GO), a layered material prepared from graphite almost 150 years ago and a precursor to chemically modified graphenes, has not been previously resolved because of the pseudo-random chemical functionalization of each layer, as well as variations in exact composition. Carbon-13 (^{13}C) solid-state nuclear magnetic resonance (SSNMR) spectra of GO for natural abundance ^{13}C have poor signal-to-noise ratios. Approximately 100% ^{13}C-labeled graphite was made and converted to ^{13}C-labeled GO, and ^{13}C SSNMR was used to reveal details of the chemical bonding network, including the chemical groups and their connections. Carbon-13–labeled graphite can be used to prepare chemically modified graphenes for ^{13}C SSNMR analysis with enhanced sensitivity and for fundamental studies of ^{13}C-labeled graphite and graphene.

Unlike crystalline materials, the structure of materials that are amorphous or that vary in chemical composition can be difficult to determine. Solid-state nuclear magnetic resonance (SSNMR) can provide important structural insights, but often requires very high enrichment of nuclei with NMR-active spins. One example of such a material that has proven difficult to characterize, despite having been first prepared almost 150 years ago (*1*), is graphite oxide (GO), which can be prepared by heating graphite in oxidizing chemicals. GO is a layered material containing interlamellar water. Materials derived from GO include its chemically functionalized (*2*), reduced (*3*), and thermally expanded forms (*4*), as well as chemically modified graphenes (*2, 3, 5–8*).

SSNMR has been done on GO but has not provided a complete understanding of the chemical structure of this material, although the detailed chemical structure has been actively researched for many years (*2, 7*). One difficulty is that the spectra do not attain a high signal-to-noise (S/N) ratio for natural abundance ^{13}C. The lack of ^{13}C-labeled GO has prevented application of modern multidimensional SSNMR methods that can provide information on the bonding arrangements of atoms and their connectivities. Although a series of one-dimensional (1D) ^{13}C SSNMR studies for GO and reduced GO revealed signal assignments and the basic chemical compositions of each, there is sparse experimental evidence of the connectivities of the chemical groups such as sp^2-bonded carbons (C=C), epoxide, carbonyl, and carboxylic groups. Thus, a variety of structural models of GO are still debated (*2*). However, we found from Monte Carlo simulations that even at only 20% ^{13}C, the abundance of ^{13}C-^{13}C bonds will be 400 times that of an unlabeled sample, so that the time required for detecting ^{13}C-^{13}C pairs in SSNMR of such a

[1]Department of Mechanical Engineering and the Texas Materials Institute, University of Texas at Austin, Austin, TX 78712, USA. [2]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China. [3]Laboratory for Space Sciences, Department of Physics, Washington University, St. Louis, MO 63130, USA. [4]Department of Chemistry, University of Illinois at Chicago, 845 West Taylor Street, Chicago, IL 60607, USA. [5]National Creative Research Initiative Center for Semiconductor Nanorods and Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang, Gyeongbuk 790-784, Korea.

*To whom correspondence should be addressed. E-mail: r.ruoff@mail.utexas.edu

Fig. 1. (A) Optical images of Ni, and SEM images of (**B**) ^{13}C-labeled synthetic graphite and (**C**) the wrinkles.

4. P. W. Shor, *Proceedings of the 35th Annual Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Los Alaminos, CA 1994), pp. 124–134.
5. C. H. Bennett, P. W. Shor, *Science* **303**, 1784 (2004).
6. B. Schumacher, M. A. Nielsen, *Phys. Rev. A* **54**, 2629 (1996).
7. W. Wootters, W. Zurek, *Nature* **299**, 802 (1982).
8. S. Lloyd, *Phys. Rev. A* **55**, 1613 (1997).
9. P. W. Shor, lecture notes from the Mathematical Sciences Research Institute Workshop on Quantum Computation, Berkeley, CA, 2002. Available online at www.msri.org/publications/ln/msri/2002/quantumcrypto/shor/1/.
10. I. Devetak, *IEEE Trans. Inf. Theory* **51**, 44 (2005).
11. D. DiVincenzo, P. W. Shor, J. A. Smolin, *Phys. Rev. A* **57**, 830 (1998).
12. G. Smith, J. A. Smolin, *Phys. Rev. Lett.* **98**, 030501 (2007).
13. C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, *Phys. Rev. Lett.* **78**, 3217 (1997).
14. M. Horodecki, P. Horodecki, R. Horodecki, *Phys. Lett. A* **223**, 1 (1996).
15. P. Horodecki, *Phys. Lett. A* **232**, 333 (1997).
16. A. Peres, *Phys. Rev. Lett.* **77**, 1413 (1996).
17. G. Smith, J. Smolin, A. Winter, *IEEE Trans. Inf. Theory* **54**, 4208 (2008).
18. K. Horodecki, M. Horodecki, P. Horodecki, J. Oppenheim, *Phys. Rev. Lett.* **94**, 160502 (2005).
19. K. Horodecki, L. Pankowski, M. Horodecki, P. Horodecki, *IEEE Trans. Inf. Theory* **54**, 2621 (2008).
20. Further details can be found in the supporting online material on *Science* Online.
21. P. Shor, J. Smolin, A. Thapliyal, *Phys. Rev. Lett.* **90**, 107901 (2003).
22. W. Dur, J. Cirac, P. Horodecki, *Phys. Rev. Lett.* **93**, 020503 (2004).
23. R. Duan, Y. Shi, *Phys. Rev. Lett.* **101**, 020501 (2008).
24. L. Czekaj, P. Horodecki, preprint available at http://arxiv.org/pdf/0807.3977.
25. P. Horodecki, M. Horodecki, R. Horodecki, *Phys. Rev. Lett.* **82**, 1056 (1999).
26. P. Shor, J. Smolin, B. Terhal, *Phys. Rev. Lett.* **86**, 2681 (2001).
27. We are indebted to C. Bennett, C. Callaway, E. Timmermans, B. Toner, and A. Winter for encouragement and comments on an earlier draft. J.Y. is supported by the CNLS and the Quantum Institute through grants provided by the Laboratory Directed Research and Development program of the U.S. Department of Energy.

Synthesis and Solid-State NMR Structural Characterization of 13C-Labeled Graphite Oxide

Weiwei Cai,[1,2] Richard D. Piner,[1] Frank J. Stadermann,[3] Sungjin Park,[1] Medhat A. Shaibat,[4] Yoshitaka Ishii,[4] Dongxing Yang,[1] Aruna Velamakanni,[1] Sung Jin An,[5] Meryl Stoller,[1] Jinho An,[1] Dongmin Chen,[2] Rodney S. Ruoff[1]*

The detailed chemical structure of graphite oxide (GO), a layered material prepared from graphite almost 150 years ago and a precursor to chemically modified graphenes, has not been previously resolved because of the pseudo-random chemical functionalization of each layer, as well as variations in exact composition. Carbon-13 (^{13}C) solid-state nuclear magnetic resonance (SSNMR) spectra of GO for natural abundance ^{13}C have poor signal-to-noise ratios. Approximately 100% ^{13}C-labeled graphite was made and converted to ^{13}C-labeled GO, and ^{13}C SSNMR was used to reveal details of the chemical bonding network, including the chemical groups and their connections. Carbon-13–labeled graphite can be used to prepare chemically modified graphenes for ^{13}C SSNMR analysis with enhanced sensitivity and for fundamental studies of ^{13}C-labeled graphite and graphene.

Unlike crystalline materials, the structure of materials that are amorphous or that vary in chemical composition can be difficult to determine. Solid-state nuclear magnetic resonance (SSNMR) can provide important structural insights, but often requires very high enrichment of nuclei with NMR-active spins. One example of such a material that has proven difficult to characterize, despite having been first prepared almost 150 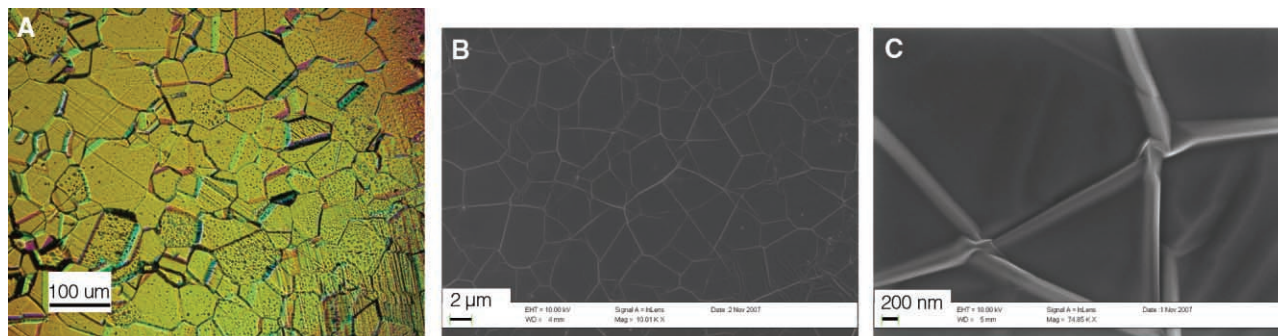years ago (*1*), is graphite oxide (GO), which can be prepared by heating graphite in oxidizing chemicals. GO is a layered material containing interlamellar water. Materials derived from GO include its chemically functionalized (*2*), reduced (*3*), and thermally expanded forms (*4*), as well as chemically modified graphenes (*2, 3, 5–8*).

SSNMR has been done on GO but has not provided a complete understanding of the chemical structure of this material, although the detailed chemical structure has been actively researched for many years (*2, 7*). One difficulty is that the spectra do not attain a high signal-to-noise (S/N) ratio for natural abundance ^{13}C. The lack of ^{13}C-labeled GO has prevented application of modern multi-dimensional SSNMR methods that can provide information on the bonding arrangements of atoms and their connectivities. Although a series of one-dimensional (1D) ^{13}C SSNMR studies for GO and reduced GO revealed signal assignments and the basic chemical compositions of each, there is sparse experimental evidence of the connectivities of the chemical groups such as sp^2-bonded carbons (C=C), epoxide, carbonyl, and carboxylic groups. Thus, a variety of structural models of GO are still debated (*2*). However, we found from Monte Carlo simulations that even at only 20% ^{13}C, the abundance of ^{13}C-^{13}C bonds will be 400 times that of an unlabeled sample, so that the time required for detecting ^{13}C-^{13}C pairs in SSNMR of such a

[1]Department of Mechanical Engineering and the Texas Materials Institute, University of Texas at Austin, Austin, TX 78712, USA. [2]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100080, China. [3]Laboratory for Space Sciences, Department of Physics, Washington University, St. Louis, MO 63130, USA. [4]Department of Chemistry, University of Illinois at Chicago, 845 West Taylor Street, Chicago, IL 60607, USA. [5]National Creative Research Initiative Center for Semiconductor Nanorods and Department of Materials Science and Engineering, Pohang University of Science and Technology, Pohang, Gyeongbuk 790-784, Korea.

*To whom correspondence should be addressed. E-mail: r.ruoff@mail.utexas.edu

Fig. 1. (A) Optical images of Ni, and SEM images of **(B)** ^{13}C-labeled synthetic graphite and **(C)** the wrinkles.

sample will be less by a factor of about 160,000 than that of unlabeled samples. Evaluation of detailed bonding networks (three neighboring ^{13}C's and so on) can be done with Monte Carlo modeling in a straightforward way for any percentage labeling of ^{13}C.

There have been a number of recent publications exploring the catalytic growth of graphite or graphitic structures (9–20). We have developed a thermal chemical vapor deposition (CVD) method for growing synthetic graphite from methane on a resistively heated Ni foil as a catalytic substrate.

A detailed description of our reactor is given in the supporting online material (21). Briefly, our reactor consists of a vacuum chamber with a metal foil supported in the center with high-current electrodes. The Ni foil can be resistively heated to near its melting point. During the deposition, the substrate temperature was held at ~1200° to 1300°C and the pressure at 1 atm. The gas used was a mixture of 10% methane and 90% Ar. The methane was 1%, 30%, 50%, 70%, and 99.95% ^{13}CH$_4$, in different growth runs. A deposition rate of ~2 μm/hour (for thin films) was obtained, and the growth rate was slowed to ~0.2 μm/hour for thicker films. Atomic force microscopy (AFM), scanning electron microscopy (SEM), x-ray diffraction (XRD), and Raman spectroscopy confirmed that the as-deposited carbon was very high-quality graphite. The graphite films deposited onto the Ni foils appeared very smooth and continuous by optical microscopy (Fig. 1). The smooth surface areas are separated from each other by wrinkles that are likely caused by the different thermal expansion coefficients of the deposited graphite and the nickel substrate (17). A typical smooth surface region is about 2 μm across, much smaller than the substrate grain sizes, and the typical wrinkle height as obtained by AFM is about 50 nm (fig. S2) (21). The ^{13}C content of the graphite samples was measured with a modified CAMECA ims3f secondary ion mass spectrometry (SIMS) instrument (CAMECA, Gennevilliers, France). The analytical error of these measurements is estimated to be ~5%, largely because the sample surfaces are uneven on a 10 to 100 μm scale. The ^{13}C content of each of the graphite films measured by SIMS was 1%, 19%, 41%, 54%, and 86% for samples prepared from 1%, 30%, 50%, 70%, and 99.95% ^{13}CH$_4$, respectively. The ^{13}C contents measured by SIMS (19%, 41%, 54%, and 86%) differed significantly from the percentages of ^{13}CH$_4$ used because for these four experimental runs the chamber had accumulated considerable residual carbon-containing material from many trial experiments with unlabeled methane. A thorough cleaning of the chamber before another run with 99.95% ^{13}CH$_4$ yielded 99.5% ^{13}C-content synthetic graphite measured by SIMS. Use of a clean chamber is thus important for the ^{13}C content of the graphite to more closely match the ^{13}C content of the input ^{13}CH$_4$ and ^{12}CH$_4$.

Raman spectra were acquired and the peak frequency shifts caused by ^{13}C-enrichment were determined. Figure 2 shows the relationship between the wave number of the G band and the ^{13}C/^{12}C ratio. The Raman frequencies shift from 1580 cm^{-1} to 1523 cm^{-1}, from unlabeled to 99.5% ^{13}C graphite (as measured by SIMS), respectively. There is good agreement between the frequency shift and the square root of the atomic mass, assuming the respective bond force constants.

AFM, SEM, and XRD data, as well as further details about SIMS and Raman analysis of ^{13}C-labeled graphite, are presented in (21).

High-resolution SSNMR using magic angle spinning (MAS) has been used as a primary method to characterize GO at the molecular level (2, 22, 23). Figure 3A shows 1D ^{13}C MAS spectrum, and Fig. 3B shows 2D ^{13}C/^{13}C chemical-shift correlation SSNMR spectrum of ^{13}C-labeled GO (made from approximately 100% ^{13}C-labeled graphite) that was prepared using a modified Hummer's method (2) with ^{13}C-labeled graphite (21). The signal assignments for the three major
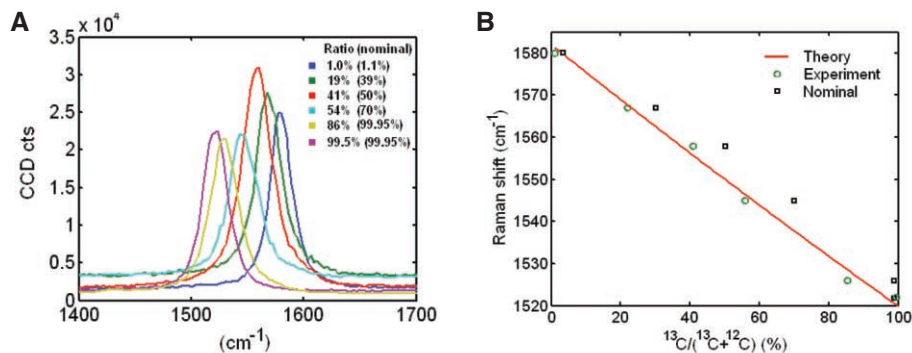
Fig. 2. (A) The G band of various ^{13}C-labeled synthetic graphites. **(B)** Shift of the G band frequency as a function of percentage of ^{13}C, as determined by SIMS. The theoretical curve was obtained as described in (21).
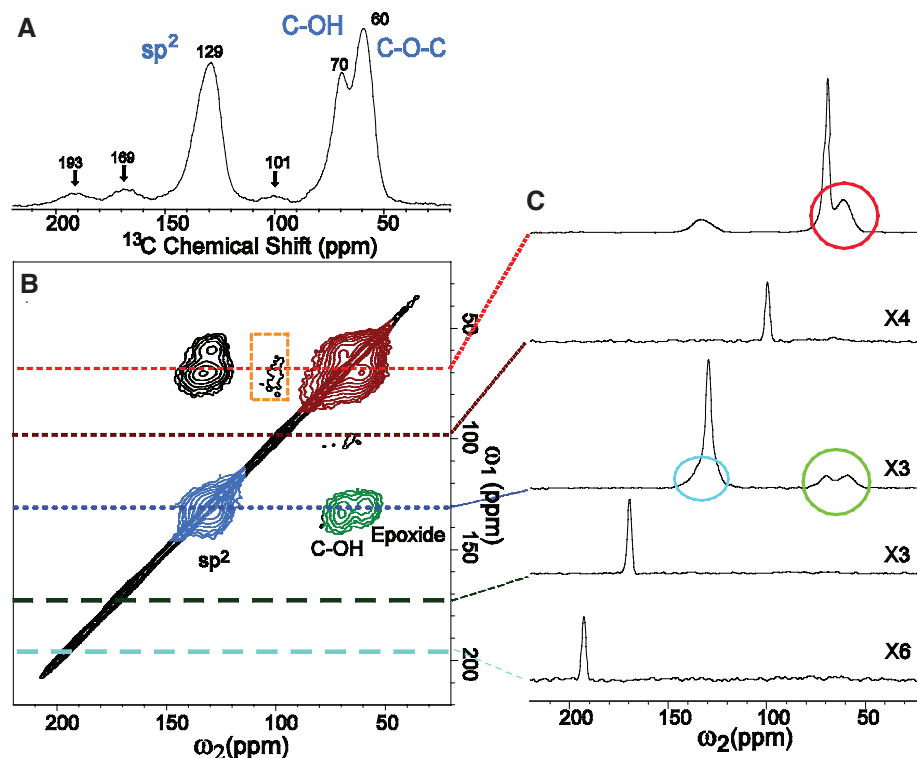
Fig. 3 (A) 1D ^{13}C MAS and **(B)** 2D ^{13}C/^{13}C chemical-shift correlation solid-state NMR spectra of ^{13}C-labeled graphite oxide with **(C)** slices selected from the 2D spectrum at the indicated positions (70, 101, 130, 169, and 193 ppm) in the ω_1 dimension. All the spectra were obtained at a ^{13}C NMR frequency of 100.643 MHz with 90 kHz ^{1}H decoupling and 20 kHz MAS for 12 mg of the sample. In (A), the ^{13}C MAS spectrum was obtained with direct ^{13}C excitation by a π/2-pulse. The recycle delay was 180 s, and the experimental time was 96 min for 32 scans. In (B), the 2D spectrum was obtained with cross polarization and fpRFDR ^{13}C-^{13}C dipolar recoupling sequence (24). The experimental time is 12.9 hours with recycle delays of 1.5 s and 64 scans for each real or imaginary t_1 point. A Gaussian broadening of 150 Hz was applied. The green, red, and blue areas in (B) and circles in (C) represent cross peaks between sp^2 and C-OH/epoxide (green), those between C-OH and epoxide (red), and those within sp^2 groups (blue), respectively.

peaks at 59.7 (epoxide ^{13}C), 69.6 (^{13}C-OH), and 129.3 (sp^2 ^{13}C) parts per million (ppm) in Fig. 3A are based on studies by Lerf *et al.* (*23*). We performed additional analyses and confirmed that these assignments are likely correct (*21*). The obtained 1D spectrum shows similar features with those reported in (*23*) except for the relatively well-resolved minor peaks at 101, 169, and 193 ppm, which respectively yield only 12%, 15%, and 4% of the integrated intensity of the 70-ppm peak. The spectrum shows a considerably stronger sp^2 peak and a much weaker peak at 169 ppm compared with that by Szabo *et al.* (*22*), although the observed peak positions are similar. The peak at 169 ppm was previously attributed to ^{13}C=O (*22*). The results imply that their sample was subject to a higher level of oxidization than ours.

Because the natural abundance of ^{13}C is only 1%, attaining sufficient sensitivity in a 2D spectrum, as shown in Fig. 3B, is extremely difficult without the ^{13}C-labeled samples. For example, ^{13}C-^{13}C bonds exist only at 0.01% abundance without labeling; thus, obtaining an equivalent 2D spectrum for an unlabeled sample would require about 10^8 times as much time. The experiment in Fig. 3B was performed with a finite-pulse radio frequency–driven dipolar recoupling (fpRFDR) mixing sequence (*24*). With the labeled sample and a relatively short mixing time (1.6 ms), the experiment permitted us to identify ^{13}C-^{13}C pairs directly bonded or separated by two bonds. In Fig. 3B, there are several strong cross peaks. For example, cross peaks were observed at the positions (ω_1, ω_2) = (133 ppm, 70 ppm) and (130 ppm, 59 ppm) (green signals in Fig. 3B). These cross peaks represent spin polarization transfer from sp^2 carbons observed at ~130 ppm in ω_1, to C-OH and epoxide groups, which appear at 70 ppm and 59 ppm in ω_2, respectively. Unlike previous studies, these cross peaks directly present the connectivity between sp^2 ^{13}C and ^{13}C-OH, as well as that between sp^2 and epoxide ^{13}C through spin-spin dipolar couplings. The cross peak intensities are about 10%, compared with the diagonal signals, which represent signals for ^{13}C spins that had the same NMR frequencies in the two dimensions ($\omega_1 = \omega_2$). The relatively strong intensities of the cross peaks suggest that a large fraction of the sp^2 ^{13}C atoms are directly bonded to ^{13}C-OH and/or epoxide ^{13}C.

We also observed strong cross peaks between ^{13}C-OH and ^{13}C-epoxide (red signals). Again, the data suggest that a large fraction of C-OH and epoxide carbons are bonded to each other. The blue cross peaks indicate that there are sp^2 species having slightly different chemical shifts and that they are bonded with each other. Indeed, the sp^2 ^{13}C shifts for the cross peaks (green) are slightly different for the cross peaks to C-OH (133 ppm) and that to epoxide (130 ppm).

In the previous studies (*23*), the proximities of the chemical groups were tentatively assigned based on formation of the phenol group during the deoxidization of GO. In contrast, the present SSNMR data directly shows that these groups are chemically bonded. For the minor species, we

found cross peaks only for the peak at 101 ppm (orange box). There are no visible cross peaks for the other minor components at 169 and 193 ppm, despite these minor peaks having comparable intensities to the 101-ppm peak. The results imply that these minor components at 169 and 193 ppm, which were previously attributed to the presence of C=O (*2*, *22*), are spatially separated from a majority of the sp^2, C-OH, and epoxide carbons.

Among six previously proposed models (*22*), only two, the Lerf-Klinowski model (*23*) and the Dékány model (*22*), present such a network. The model proposed by Dékány *et al.* may be correct for their more highly oxidized compound, because that structural model seems to call for a considerably higher level of oxidization to complete the modification of an sp^2 network into a network of linked cyclohexanes. Further studies would be needed to define all of the structural details of the system.

Chemically modified graphenes that will be of importance in a variety of new materials can now be ^{13}C-labeled and more effectively studied by SSNMR. High-quality ^{13}C-labeled graphite should find use for fundamental property measurements, including of ^{13}C-labeled graphene.

References and Notes
1. B. C. Brodie, *Annales des Chimie et des Physique* **59**, 466 (1860).
2. S. Stankovich *et al.*, *Carbon* **45**, 1558 (2007).
3. S. Stankovich *et al.*, *J. Mater. Chem.* **16**, 155 (2006).
4. M. J. McAllister *et al.*, *Chem. Mater.* **19**, 4396 (2007).
5. D. A. Dikin *et al.*, *Nature* **448**, 457 (2007).
6. S. Stankovich *et al.*, *Nature* **442**, 282 (2006).
7. S. Stankovich, R. D. Piner, S. T. Nguyen, R. S. Ruoff, *Carbon* **44**, 3342 (2006).
8. S. Watcharotone *et al.*, *Nano Lett.* **7**, 1888 (2007).
9. H. H. Angermann, G. Horz, *Appl. Surf. Sci.* **70-71**, 163 (1993).
10. R. Anton, *J. Mater. Res.* **20**, 1837 (2005).
11. A. Barbangelo, R. Sangiorgi, *Mater. Sci. Eng. A* **156**, 217 (1992).
12. F. Bonnet, F. Ropital, Y. Berthier, P. Marcus, *Mater. Corrosion Werkst. Korros.* **54**, 870 (2003).
13. D. V. Fedoseev, S. P. Vnukov, B. V. Derjaguin, *Carbon* **17**, 453 (1979).
14. D. Fujita, T. Homma, *Surf. Interface Anal.* **19**, 430 (1992).
15. D. Fujita, K. Yoshihara, *J. Vac. Sci. Technol. A* **12**, 2134 (1994).
16. C. Klink, I. Stensgaard, F. Besenbacher, E. Laegsgaard, *Surf. Sci.* **342**, 250 (1995).
17. A. N. Obraztsov, E. A. Obraztsova, A. V. Tyurnina, A. A. Zolotukhin, *Carbon* **45**, 2017 (2007).
18. H. Oudghiri-Hassani, S. Rakass, N. Abatzoglou, P. Rowntree, *J. Power Sources* **171**, 850 (2007).
19. R. Sinclair, T. Itoh, R. Chin, *Microsc. Microanal.* **8**, 288 (2002).
20. M. Yudasaka, R. Kikuchi, Y. Ohki, S. Yoshimura, *J. Vac. Sci. Technol. A* **16**, 2463 (1998).
21. Materials and methods are available as supporting material on *Science* Online.
22. T. Szabo *et al.*, *Chem. Mater.* **18**, 2740 (2006).
23. A. Lerf, H. Y. He, M. Forster, J. Klinowski, *J. Phys. Chem. B* **102**, 4477 (1998).
24. Y. Ishii, *J. Chem. Phys.* **114**, 8473 (2001).
25. ^{13}CH$_4$ was provided through a research grant from Cambridge Isotopes Laboratories, Inc. We (R.S.R.) wish to acknowledge The University of Texas for startup funds and for partial support by the *Defense Advanced Research Projects Agency* Center on Nanoscale Science and Technology for Integrated Micro/NanoElectromechanical Transducers (HR0011-06-1-0048). The WiTec Micro-Raman instrument was acquired by an Air Force Office of Scientific Research, Defense University Research Instrumentation Program grant. Y.I. appreciates support from the Dreyfus Foundation Teacher-Scholar Award program and the NSF CAREER program (CHE 449952). S.J.A. was financially supported by the Korea Science and Engineering Foundation under the National Creative Research Initiative Project (R16-2004-004-01001-0) of the Ministry of Science and Technology, Korea. J. Goodenough and A. Ruoff commented on an early version of this manuscript.

Linear Response Breakdown in Solvation Dynamics Induced by Atomic Electron-Transfer Reactions

Arthur E. Bragg, Molly C. Cavanagh, Benjamin J. Schwartz*

The linear response (LR) approximation, which predicts identical relaxation rates from all nonequilibrium initial conditions that relax to the same equilibrium state, underlies dominant models of how solvation influences chemical reactivity. We experimentally tested the validity of LR for the solvation that accompanies partial electron transfer to and from a monatomic solute in solution. We photochemically prepared the species with stoichiometry Na^0 in liquid tetrahydrofuran by both adding an electron to Na$^+$ and removing an electron from Na$^-$. Because atoms lack nuclear degrees of freedom, ultrafast changes in the Na^0 absorption spectrum reflected the solvation that began from our two initial nonequilibrium conditions. We found that the solvation of Na^0 occurs more rapidly from Na$^+$ than Na$^-$, constituting a breakdown of LR. This indicates that Marcus theory would fail to describe electron-transfer processes for this and related chemical systems.

Solvent-solute interactions play an integral role in solution-phase chemical reactivity and particularly in electron-transfer (ET) reactions (*1*), in which solvation dynamics—the

response of the solvent to changes in solute size and/or electronic charge distribution (*2*)—help drive the motion of charge from donor to acceptor. Current theoretical understanding of how

solvation dynamics influence chemical reactivity [for example, through the Marcus theory of ET (*1*)] is based largely on the idea of linear response (LR) (*3–6*). If the Hamiltonian governing solute-solvent interactions is assumed to change linearly with respect to a perturbation that moves the system out of equilibrium, then the relaxation of the perturbed nonequilibrated state should be identical to the regression of the spontaneous fluctuations of the system at equilibrium; this assumption is commonly called the LR approximation (*3*). In this work, we performed an experimental test of LR in which we investigated the solvation dynamics associated with elementary steps that play a role in many common chemical transformations, specifically the spontaneous partial transfer of an electron to and from an atomic solute in solution. We demonstrate that the LR approximation fails to underlie the solvation processes associated with this particular atomic ET reaction, leading us to predict that LR may break down more broadly in solution-phase chemical dynamics than previously appreciated.

A convenient way to characterize the evolution of solute-solvent interactions is to use the framework of time correlation functions (TCFs). At equilibrium, the evolution of solvent-solute interactions can be quantified by Eq. 1

$$C(t) = \frac{\langle \delta E(t) \delta E(0) \rangle}{\langle \delta E(0)^2 \rangle} \qquad (1)$$

which autocorrelates the fluctuations of the solute-solvent interaction energy E about its average value, $\delta E(t) = E(t) - \langle E \rangle$; the angled brackets represent equilibrium ensemble averages. This TCF can be viewed as describing the time scale of the intrinsic fluctuations of an equilibrium solvent-solute system: if we artificially prepared a specific solvent-solute configuration that could be reached via thermal fluctuations at equilibrium, then $C(t)$ would reflect the time dependence of the subsequent relaxation away from this configuration. Within the limit of LR, the nonequilibrium relaxation of the solvent-solute interaction energy after perturbation of the system from equilibrium, $S(t)$, should match the TCF of the system at equilibrium (*3, 7, 8*)

$$S(t) = \frac{\overline{E}(t) - \overline{E}(\infty)}{\overline{E}(0) - \overline{E}(\infty)} \approx C(t) \qquad (2)$$

where the overbars indicate nonequilibrium ensemble averages. Although the LR approximation (Eq. 2) can be derived from statistical mechanical perturbation theory {and thus is applicable when the perturbation to a system is near thermal energies [Boltzmann's constant

Department of Chemistry and Biochemistry, University of California, Los Angeles Los Angeles, CA 90095-1569, USA.

*To whom correspondence should be addressed. E-mail: schwartz@chem.ucla.edu

times the temperature $(\sim k_B T)$] (*3*)}, it has recently been shown that LR arises whenever the fluctuations of a system are Gaussian, regardless of the magnitude of the deviation from equilibrium (*5, 7, 8*). It is these Gaussian statistics, in turn, that give rise to the familiar parabolic potentials along the generalized solvent coordinate that lie at the foundation of the Marcus theory of ET (*1*). Thus, when valid, the LR approximation provides a useful means of estimating and interpreting theoretically the time-dependent relaxation that follows the preparation of many (*1, 3–5, 8*), but not all (*6, 7, 9*), highly perturbed solute-solvent configurations.

How can we test the validity of LR experimentally, particularly for the solvation dynamics tied to chemical reactivity? To test LR directly, we must not only track the dynamics of quantity associated with the time-evolving solute-solvent interaction as the system relaxes from a well-defined nonequilibrium configuration [$S(t)$], but also measure the solvent-solute fluctuations at equilibrium [$C(t)$]. Unfortunately, the equilibrium TCF is difficult to access, although a recently developed approach that mea-

sures the difference in solvent spectral density after the photoexcitation of a probe solute shows great promise toward making this direct comparison (*10*). Here, we describe an alternative approach to assess the LR approximation: We compare the relaxation dynamics that begin from two different well-defined nonequilibrium configurations but end in the same final equilibrium state. If LR holds, then the time dependence of the solvent-solute relaxation from both nonequilibrium states should be identical to the equilibrium TCF. If the two nonequilibrium configurations relax on different time scales, as we discuss below, then the LR prediction must fail to describe one or both of the relaxation pathways.

Although our approach should apply to any solute-solvent system, we have chosen to survey the relaxation of atoms in solution because these solutes lack internal nuclear degrees of freedom. Thus, unlike the large molecular chromophores that are typically used as solvation probes, the atomic relaxation dynamics that we observe directly reflect the motions of the surrounding solvent and are unobscured by competing intramolecular processes (*11*). Specifically, we ex-

Fig. 1. (A to C) Illustration of how LR can be tested by examining the dynamic solvation of the same solute prepared from two different well-defined initial conditions. **(A)** Schematic depiction of the local solvent reconfiguration required to solvate the nonequilibrium (Na⁺, e⁻) TCP (green circles labeled 0) prepared after both electron detachment from Na⁻ (right, blue circle with minus sign) and electron attachment to Na⁺ (left, red circle with plus sign). In both cases, the solvent must relax to accommodate changes in solute size and charge that result from partial ET reactions. **(B)** Another representation of the two solvent relaxation pathways shown in (A), presented using the picture derived from the Marcus theory of ET. **(C)** A schematic presentation of the time dependence of the two nonequilibrium solvent relaxation processes shown in (A)

and (B); the ordinate in both (B) and (C) is the solute-solvent interaction energy. **(D)** The steady-state equilibrium spectra of the chemical species examined in this work: the absorption spectra of the equilibrated THF-solvated (Na⁺, e⁻) TCP [green solid curve (*18*)]; the THF-solvated electron [e^-_{THF}, purple dashed curve (*19*)]; the THF-solvated Na⁻ [blue dashed curve (*23*)]; and the CTTS transition of the Na⁺-I⁻ ion pair in THF (red dash-dotted curve). We used both Na⁻ and Na⁺-I⁻ as precursors for the photochemical production of nonequilibrium (Na⁺, e⁻) TCPs in liquid THF via electron detachment and attachment, respectively.

amined the nonequilibrium relaxation of neutral atomic sodium (Na^0) in liquid tetrahydrofuran (THF) that was prepared both by photoinduced ET to sodium cations (Na$^+$) and by electron photodetachment from sodium anions (Na$^-$, or sodide). Figure 1A schematically illustrates the solvent reconfiguration that occurs after each of these photochemical preparations; Fig. 1B presents this same relaxation in terms of the picture derived from the Marcus theory of ET, in which the relaxation of the solvent-solute interaction energy, E, is expressed as a function of a collective solvent coordinate; and Fig. 1C illustrates the time-dependent solute-solvent interaction energy as would be typically examined in a molecular dynamics (MD) simulation. The red and blue curves in Fig. 1C reflect fluctuations of the ion-solvent interactions about their equilibrium average values before each photochemical process. If we promptly add or remove an electron to or from one of the initially equilibrated ions through photoinduced ET (vertical black arrows in Fig. 1, A to C), the solvent interaction energy associated with the newly created neutral solute will be outside the range of the equilibrium fluctuations [green curves labeled $C(t)$ in Fig. 1, B and C], because the initial solvent structure is equili-

brated for an ion and thus poorly accommodates the neutral Na. The nearby solvent molecules will then relax by both translating to accommodate the change in solute size and rotating so that the solvent dipoles become favorably oriented for the neutral Na. We label the nonequilibrium relaxation of the neutral Na that begins with the initial configuration characteristic of the Na anion "$S_-(t)$" (right side of Fig. 1, A to C), and we label the relaxation that begins from the initial configuration associated with the Na cation "$S_+(t)$" (left side of Fig. 1, A to C). If LR holds, then we expect $S_+(t) = S_-(t)$ because both should be equal to $C(t)$.

Atomic Na in THF is a convenient system for testing the LR approximation through this approach for several reasons. The precursors we used to create neutral Na photochemically (Na$^-$ and the Na$^+$-I$^-$ ion pair) are stable in liquid THF. In addition, the solvent structure and the nature of the delocalized solvent-supported electronic states of liquid THF (12, 13) allow us to rapidly shuttle electrons to and from these ion precursors via photoinduced ET (14, 15). Finally, the spectrum of atomic Na in THF is highly sensitive to the nature of the local solvent environment (16) and spans visible and near-infrared wavelengths (Fig. 1D, green solid curve),

so we can easily track the solvation dynamics that follow each photochemical preparation, using transient absorption spectroscopy.

The extraordinary sensitivity of the Na atom's absorption spectrum to the local solvent environment arises from the chemical nature of Na in THF: Rather than a solvated atom, the equilibrated neutral Na solute in THF is better thought of as a (Na$^+$, e^-) tight-contact pair (TCP), in which the electron has substantial interactions with the solvent as well as a partial-valence interaction with the Na cation core (16, 17). The steady-state absorption spectrum of the (Na$^+$, e^-) TCP peaks at 870 nm [Fig. 1D, green solid curve (18)], a wavelength that lies between the narrow gas-phase Na D line (peaking at 590 nm) and the broad absorption of the THF-solvated electron, e^-_{THF} [Fig. 1D, purple dotted curve, peaking at 2160 nm (19)], reflecting the fact that the chemical nature of this solute lies between the extremes of pure cation-electron and solvent-electron interactions. As a result, the TCP is formed from either a solvent-separated Na$^+$ + e^-_{THF} pair or a weakly solvated, gas-phase–like Na^0 atom by partial transfer of an electron toward or away from the Na$^+$ core, respectively

$$(\mathrm{Na^+ \cdots} e^-) \xrightarrow{\text{partial ET}} (\mathrm{Na^+},\ e^-) \xleftarrow{\text{partial ET}} \mathrm{Na^0} \quad (3)$$

In our experiments, we used two different photoinduced ET processes (Fig. 1, A to C) to create the species on both the left and right of Eq. 3 and then track the solvation dynamics associated with the spontaneous partial ET reactions that generate (Na$^+$, e^-) as a result of the subsequent solvent-solute relaxation.

The THF-solvated (Na$^+$, e^-) TCP can be generated readily from its consituent parts, Na$^+$ and e^- (left side of Eq. 3) (18), but under most conditions it is challenging to measure the dynamic solvation of a solute produced through electron attachment, because the rate of attachment is generally limited by the mutual diffusion of the reactants. However, by taking advantage of the relatively strong ion-pair interactions that exist in weakly polar solvents such as THF, we can locate an electron-donating chromophore (here, iodide) close to the Na cation and thus photo-initiate prompt electron attachment. In previous work, we showed that electrons generated through charge-transfer–to–solvent (CTTS) excitation of counterion-free iodide in THF were ejected ~6 nm away from the iodine core (20), but that the presence of nearby Na$^+$ collapses the CTTS-electron ejection distribution within ~2 nm of the cation, so that >50% of the electrons attach to Na$^+$ to form (Na$^+$, e^-) TCPs within 2 ps (21). Consequently, the Na$^+$-I$^-$ ion pair in THF provides an ideal precursor to create our (Na$^+$, e^-) TCP solvation probe from its constituent parts via excitation of the Na$^+$-I$^-$ CTTS band (Fig. 1D, red dot-dashed curve) and to subsequently probe in detail the ultrafast solvation dynamics associated with electron attachment, $S_+(t)$.

Fig. 2. Ultrafast transient absorption dynamics after the 263-nm CTTS excitation of Na$^+$-I$^-$ in liquid THF. (**A**) Spectral reconstruction (SOM text) showing the dynamics of Na$^+$-e^- attachment to form the (Na$^+$, e^-) TCP; the symbols represent the data, which are connected with colored lines to guide the eye (the large linear region near 800 nm results from difficulty in probing near the laser fundamental with our transient absorption setup; error bars represent 95% confidence limits of measurements). The fact that the early-time data do not match the equilibrium spectrum of e^-_{THF} (shown as the long-dashed black curves that have been scaled for ease of comparison to the data at 1.5 and 12 ps) near ~1700 nm indicates that some of the CTTS-generated electrons rapidly attach to Na$^+$ to form Na cation-electron LCPs. The short-dashed black curve shows the equilibrium (Na$^+$, e^-) TCP spectrum for reference. (**B**) Spectral dynamics associated with the (Na$^+$, e^-) TCP created after electron attachment [colored lines, as in (A)]; the spectrum of this species has been isolated by subtracting the contributions of e^-_{THF} and (Na$^+ \cdots e^-$)$_{LCP}$ (SOM text). The data show that after the LCP → TCP partial ET reaction, dynamic solvation results in a spectral blue shift of the nonequilibrium TCP absorption that occurs on a ~5-ps time scale. Different wavelength scales are used in (A) and (B).
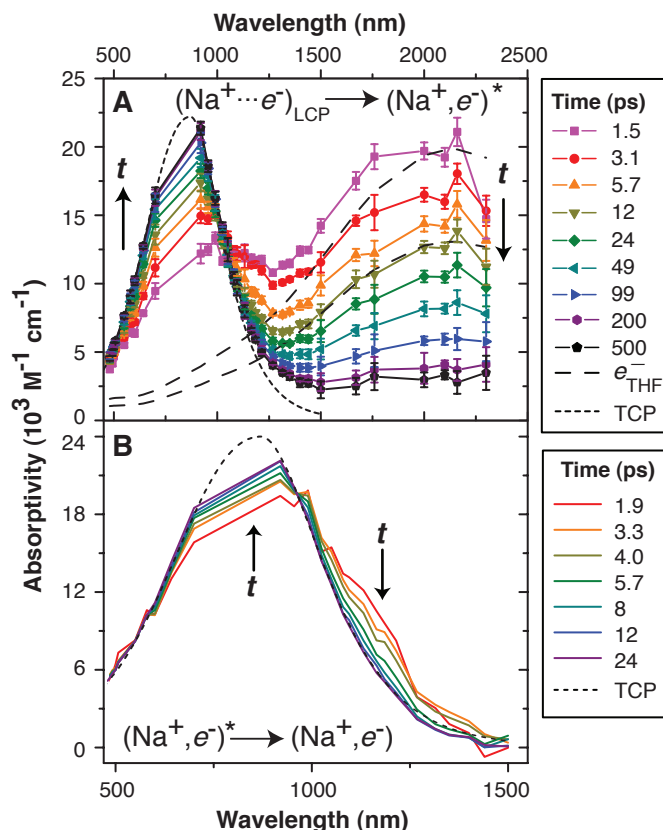
Figure 2A presents the temporal evolution of the transient absorption that follows the 263-nm CTTS excitation of 20-mM NaI in THF [see the supporting online material (SOM) text for details (22)]. The data show that CTTS excitation rapidly induces an absorption in the 1200- to 2300-nm spectral region. This absorption then decays over the course of 200 to 300 ps concurrently with the appearance of a new absorption band in the 480- to 900-nm region. The absorption dynamics at intermediate wavelengths (900 to 1200 nm) evolve with more complexity, although a quasi-isosbestic point appears near 1130 nm. To first order, these spectral dynamics reflect the disappearance of CTTS-generated THF-solvated electrons (e^-_{THF}) as they diffusively attach to nearby Na$^+$ in solution to form the (Na$^+$, e^-) TCP (21).

A closer examination of Fig. 2A, however, reveals that the spectrum measured between 1200 and 2300 nm immediately after CTTS excitation of NaI (at 1.5 ps, pink trace) does not match that of the equilibrated THF-solvated electron (scaled and plotted as long-dashed curves). We have observed a similar spectrum (that is, slightly blue-shifted relative to the spectrum of e^-_{THF}) after the CTTS excitation of tetrabutyl-ammonium iodide (t–BA$^+$-I$^-$) in THF (21), and we determined that the weak ion-pair interactions between I$^-$ and t-BA$^+$ promote the forma-

tion of a loose cation-electron contact pair (LCP). The LCP is characterized by a weak Coulombic cation-electron attraction (not the stronger partial valence interactions that define the TCP), and its spectrum can be thought of as arising from a Stark shift of the e^-_{THF} absorption induced by the presence of the nearby cation. For the Na$^+$-I$^-$ system studied here, the spectrum between 1200 and 2300 nm (Fig. 2A) rapidily decays in intensity over the initial few picoseconds as TCPs are formed, with the remaining feature matching the shape of the e^-_{THF} spectrum by ~12 ps (light green trace). Thus, the data in Fig. 2A demonstrate the transient formation of a sizeable Na$^+$-e^- LCP population after CTTS excitation of Na$^+$-I$^-$ in THF (21, 22)

$$I^- + Na^+ \xrightarrow{hv, CTTS} I + (Na^+ \cdots e^-)_{LCP} \quad (\sim 60\%)$$

$$\text{or } I + Na^+ + e^-_{THF} \quad (\sim 40\%) \quad (4a)$$

$$(Na^+ \cdots e^-)_{LCP} \xrightarrow[2.3 \text{ ps}]{\text{partial ET}} (Na^+, e^-)^* \xrightarrow{S_+(t)}$$

$$(Na^+, e^-) \quad (4b)$$

$$Na^+ + e^-_{THF} \xrightarrow[63 \text{ ps}]{\text{diffuse}} (Na^+ \cdots e^-)_{LCP} \xrightarrow{\text{via eq. 4b}}$$

$$(Na^+, e^-) \quad (4c)$$

(Na$^+$, e^-)* denotes a nonequilibrated TCP solute, and I is a spectator in Eqs. 4b and 4c [(SOM text) (22)]. As a result, CTTS excitation of NaI allows us to rapidly create a population of LCPs, which then spontaneously convert to a nonequilibrium (Na$^+$, e^-)* TCP population via partial ET to Na$^+$.

We can isolate the spectral evolution associated with TCP solvation by subtracting the spectral contributions from the LCP and e^-_{THF} [the latter formed in ~40% yield after CTTS ejection from iodide to locations far from a Na$^+$ cation (Eq. 4a)], and by normalizing for TCP population kinetics, as plotted in Fig. 2B [(SOM text) (22)]. Although the first ~2 ps of the TCP spectral dynamics are obscured by the LCP → TCP interconversion process as well as by experimental resolution effects (21), the data clearly show that the initially produced (Na$^+$, e^-)* species (at 1.9 ps) has a spectrum that is red-shifted and slightly broadened relative to the equilibrium absorption (black dashed curve). As the solvent relaxes, the transient spectrum of the TCP blue-shifts and narrows on a ~5-ps time scale, with solvation complete by 10 ps. Because the TCP solvation dynamics we measured may be rate-limited by the LCP → TCP reaction process (the first step of Eq. 4b), this ~5-ps time scale represents an upper limit for the decay of $S_+(t)$.

To examine the nonequilibrium solvation of the same (Na$^+$, e^-) TCP solute from a different initial solvent-solute configuration (Fig. 1, A to C, right panels), we also created the TCP in a solvent environment that initially accommodates its corresponding anion, Na$^-$. Like I$^-$, Na$^-$ in liquid THF has a strong CTTS transition [Fig. 1D, blue dashed curve (23)], photoexcitation of which leads to subpicosecond electron ejection and

Fig. 3. Spectrally reconstructed ultrafast transient absorption dynamics of the neutral Na species created after 395-nm CTTS excitation of Na$^-$ in liquid THF [(16) and SOM text (22)]. (**A**) Spectral dynamics (squares connected by colored lines to guide the eye; error bars represent 95% confidence limits of measurements) associated with the chemical conversion of a weakly solvated Na atom, characterized by the split and broadened Na D-line transition near 600 nm (0.38 ps), to become the THF-solvated (Na$^+$, e^-) TCP; the isosbestic point near 720 nm indicates that Na^0 and (Na$^+$, e^-) TCPs are chemically distinct species that interconvert via a spontaneous partial ET reaction. (**B**) Spectral dynamics associated with the (Na$^+$, e^-) TCP after electron detachment (colored lines); the spectral dynamics of this species have been isolated by subtracting the contribution from Na^0 and have been normalized to remove contributions from changes in TCP population [(16) and SOM text]. The data demonstrate that after the Na^0 → (Na$^+$, e^-) partial ET reaction, dynamic solvation results in a substantial red shift of the nonequilibrium TCP absorption that occurs on a ~10-ps time scale.
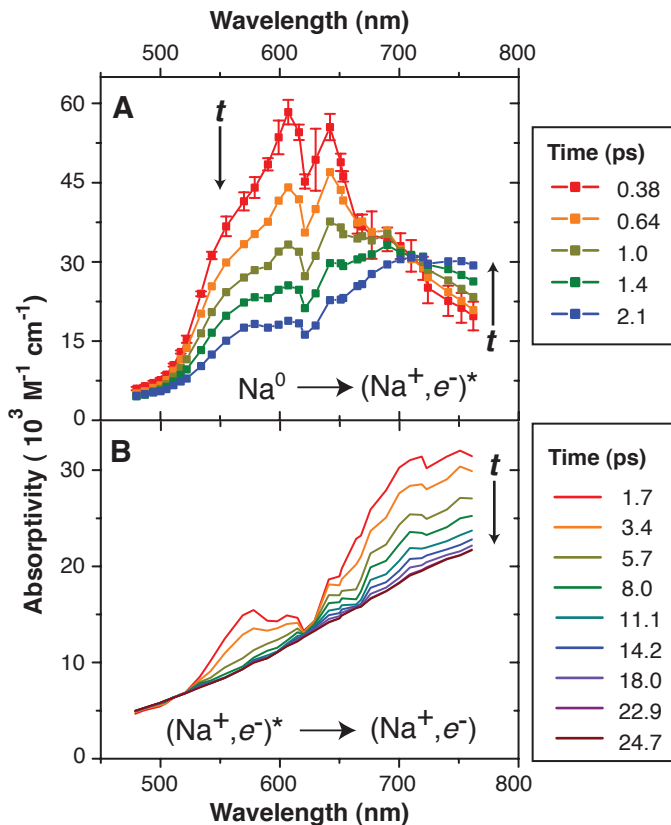
Fig. 4. Ultrafast spectral dynamics, probed at 1130 nm, of the (Na$^+$, e^-) TCP created after both electron attachment via 263-nm CTTS excitation of Na$^+$-I$^-$ [circles, $S_+(t)$] and electron detachment via 395-nm CTTS excitation of Na$^-$ [squares, $S_-(t)$]. The two spectral transients have been normalized at 2 ps and the detachment data have been inverted for ease of comparison. To the extent that probing at a single wavelength characterizes the entire solvent response (26), the data show that solvation of the nonequilibrium TCP occurs roughly twice as fast after electron attachment than after electron detachment (1/e time scales of 3.9 and 7.5 ps, respectively), thus demonstrating a breakdown of LR.

leaves behind a neutral atomic Na core. Because of the cavities that are inherent in the structure of liquid THF (12, 13), the distance to which the CTTS-generated electrons are ejected from the neutral Na core increases with increasing excitation energy (24). Thus, by exciting the Na⁻ CTTS transition at 395 nm, we are able to move the CTTS-generated electron far from the Na⁰ core, so that the spectral dynamics we measure for Na⁰ are negligibly affected by the proximity of (or possible recombination with) the CTTS-ejected electron (16). The spectral dynamics associated with the Na⁰ product generated via 395-nm excitation of the Na⁻ CTTS band, corrected for the modest Na⁰-e^-_{THF} recombination that does take place, are shown in Fig. 3 (16, 22).

The spectrum of the nascent neutral Na atom measured immediately after electron photodetachment from Na⁻ (at 0.38 ps) exhibits a broad, intense absorption band that peaks between 600 and 650 nm. We (16) and others (25) have assigned this feature to the sodium D-line transition, albeit broadened and split by the asymmetry of the surrounding solvent cavity. The similarity in the position of this band to that measured in the gas phase reflects the weak interaction of the nascent neutral atom with the surrounding solvent; this is expected because the initial solvent structure had accommodated equilibrated Na⁻, which is much larger than the neutral Na atom (Fig. 1A, right) (16). Figure 3A further reveals that this initially produced D-line absorption feature decays substantially over the next ~2 ps and that the decay is accompanied by a concomitant increase in spectral intensity at longer wavelengths (>700 nm); a well-defined isosbestic point at ~720 nm indicates that these spectral changes result from a kinetic interconversion between two chemically distinct species. Careful analysis of this spectral progression reveals that the neutral Na solute starts from a state in which the 3s valence electron is bound almost exclusively by the Na⁺ core but then undergoes a spontaneous partial ET reaction on a ~730-fs time scale to produce a nonequilibrated (Na⁺, e^-)* TCP in which the electron is subject to substantial interactions with the solvent (16)

$$\text{Na}^- \xrightarrow{h\nu,\ CTTS} \text{Na}^0 + e^-_{THF} \qquad (5a)$$

$$\text{Na}^0 \xrightarrow[730\ fs]{partial\ ET} (\text{Na}^+,\ e^-)^* \xrightarrow{S_-(t)} (\text{Na}^+,\ e^-) \quad (5b)$$
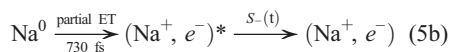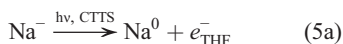
After this partial ET from the nascent neutral atom (which was produced via photoinduced CTTS detachment from Na⁻), (Na⁺, e^-)* relaxes with the surrounding solvent to reach equilibrium, $S_-(t)$.

Figure 3B highlights the spectral evolution of the TCP during the $S_-(t)$ relaxation, which has been isolated by removing contributions from the Na⁰ absorption at 600 nm and normalized for its time-dependent population kinetics (16, 22). The salvation induced evolution of the (Na⁺, e^-)

spectrum occurs more slowly than the partial ET reaction that forms this species from Na⁰. Figure 3B shows that the time-dependent TCP spectrum red shifts (and broadens to maintain a constant oscillator strength) as it relaxes to equilibrium. The data show that this dynamic solvation process takes place on a ~10-ps time scale and that the $S_-(t)$ relaxation is not complete for at least 20 to 25 ps.

One way to approximately quantify the $S_+(t)$ and $S_-(t)$ TCFs inherent to the data in Figs. 2B and 3B is to compare the dynamics at a single wavelength (26). Thus, in Fig. 4, we compare the time-dependent TCP spectral dynamics at 1130 nm for both solvation processes. We chose this wavelength for two reasons. First, 1130 nm lies near the isosbestic point for the TCP and LCP/e^-_{THF} absorption bands; thus, for the photo-induced electron-attachment pathway, this wavelength probes primarily TCP solvation and not population dynamics, guaranteeing that the kinetic model we used to deconstruct our data does not influence the $S_+(t)$ solvation response we derive at this wavelength [(SOM text) (22)]. Second, the long-time TCP solvation after electron photodetachment is observed most cleanly at this wavelength, because there is no interference from Na⁻ bleach and the e^-_{THF} absorption contribution can be removed readily (16). To directly compare the 1130-nm spectral dynamics from the two pathways, we have inverted the $S_-(t)$ response from the photodetachment pathway, because the TCP spectral red shift results in a delayed rise (rather than a decay) at this wavelength (16). We also have normalized the 1130-nm spectral transients at 2 ps in order to focus on the long-time solvation dynamics and to exclude contributions from the faster ET interconversion kinetics observed along both relaxation pathways. Figure 4 indeed illustrates that, as estimated by the 1130-nm spectral dynamics, $S_+(t)$ decays roughly twice as fast as $S_-(t)$ (1/e time scales of 3.9- and 7.5-ps, respectively). Because we know that the $S_+(t)$ response is partly rate-limited by the LCP → TCP conversion process, whereas the $S_-(t)$ response occurs more slowly than the Na⁰ → (Na⁺, e^-) reaction time scale, this difference in the two responses represents a lower limit.

Together, Figs. 2B, 3B, and 4 prove that LR does not apply for the solvation dynamics that accompany the creation of the (Na⁺, e^-) TCP through the two different partial ET reactions. The $S_+(t)$ solvation response that follows the creation of the TCP via partial ET from a LCP [which was created via electron photoattachment (Eq. 4)] starts from a solvent configuration that was optimized to accommodate the small Na⁺, prompting a blue shift of the TCP spectrum that is complete within ~10 ps; this time scale is similar to the solvent relaxation measured after the excitation of a dye molecule in liquid THF (27). In contrast, the $S_-(t)$ solvation response that follows the creation of the TCP via partial ET from Na⁰ (Eq. 5) starts from a solvent configuration

that was optimized to accommodate the large Na⁻ anion, prompting a red shift of the TCP spectrum that is not complete for 20 to 25 ps. This large difference in time scale in the solvation dynamics of the same species approaching equilibrium from two different initial configurations is a clear indication that the LR approximation, which predicts identical $S_+(t)$ and $S_-(t)$ responses, does not describe the solvation dynamics that follow these simple partial electron-transfer reactions.

Why should we observe this breakdown of LR? Guided by results of MD simulations that have demonstrated a failure of LR similar to what we observed here (6), we believe that the changes in solute size involved in these partial charge-transfer reactions help promote the breakdown of LR. These simulations revealed that when the solute size decreases (as occurs upon the removal of an electron), the rate of the solvent response is essentially limited by the translational diffusion of first-shell solvent molecules into the void space left behind after the solute shrinks. Thus, we expect LR to break down in this case both because the initial solvent configuration about the smaller solute is never explored through equilibrium fluctuations and because the corresponding diffusional solvent motions that relax this configuration are slower than those at equilibrium. In contrast, the simulations also have shown that the solvent response that accompanies a solute size increase (as occurs upon the addition of an electron) is driven rapidly by strong, short-ranged repulsive interactions between the enlarged solute and the first-shell solvent molecules. The fact that $S_-(t)$ is substantially slower than $S_+(t)$ strongly suggests that the relative changes in size of the initial Na⁰ and LCP as they become the TCP are primarily responsible for the breakdown of LR that we observed in our experiments.

To what extent do the processes we have studied reflect the solute-solvent relaxation in more common chemical transformations? Is the breakdown of LR we observed in this model system the rule, and not an exception, for the nonequilibrium solvation that occurs in the course of typical chemical processes? In addition to our observations, a breakdown of LR has recently been observed to follow the photodissociation of the ICN molecule in ethanol (28), which produces a CN fragment whose excited rotational motion is only slowly quenched by interactions with the surrounding solvent. This behavior demonstrates an anticipated failure of LR in scenarios in which a chemical process partitions substantial internal energy (many k_BT) into a product fragment (7, 28). Even though the photochemical preparation in our experiments involves photons with energies of hundreds of k_BT, most of this energy is used to detach an electron from either Na⁻ or I⁻ and move it through liquid THF; the bulk of this energy is not partitioned directly to the TCP product. Moreover, our experiments measure the TCP solvation only after spontaneous partial ET

reactions (Eqs. 3 to 5). Thus, although we may anticipate that the solvation dynamics measured involve substantial changes in the local solvation structure due to the solute size change, we expect that the energies associated with $(Na^+, e^-)^*$ formation are not chemically extreme and are representative of those associated with common solution-phase reactions.

This clear breakdown of LR implies that the solvent fluctuations coupled to the (Na^+, e^-) TCP are not Gaussian, and thus that the potential surfaces associated with these ET processes are highly nonparabolic. As a result, the Marcus theory of ET would poorly describe these ET processes. We anticipate that this could be an important consideration in many similar outer-sphere ET reactions, where substantial rearrangement of the local solvent structure could induce a similar LR breakdown. It is also important to note that the LR approximation is built on the idea that the same solvent-solute motions that underlie equilibrium fluctuations are also responsible for the nonequilibrium solvation dynamics (3). Although we observed a clear difference in the time dependence of two solvation pathways that reflects a breakdown of LR, observing an identical time dependence would not have guaranteed that the LR holds. This is because even when the specific molecular motions responsible for relaxing a nonequilibrium perturbation differ considerably from the solvent fluctuations active at equilibrium, LR may appear to be valid if the relevant nonequilibrium and equilibrium solvent

motions happen to occur on similar time scales; what we have termed a hidden breakdown of LR (29, 30). Overall, these findings demonstrate that an accurate assessment of solvation dynamics—and, by extension, our understanding of solution-phase chemical reactivity—must be considered directly at the molecular level in order to determine correctly how best to understand the solvent relaxation resulting from a given nonequilibrium perturbation.

References and Notes

1. R. A. Marcus, N. Sutin, *Biochim. Biophys. Acta* **811**, 265 (1985).
2. M. Maroncelli, *J. Mol. Liq.* **57**, 1 (1993).
3. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford Univ. Press, New York, 1987).
4. R. M. Stratt, M. Maroncelli, *J. Phys. Chem.* **100**, 12981 (1996).
5. P. L. Geissler, D. Chandler, *J. Chem. Phys.* **113**, 9759 (2000).
6. D. Aherne, V. Tran, B. J. Schwartz, *J. Phys. Chem. B* **104**, 5382 (2000).
7. G. Tao, R. M. Stratt, *J. Chem. Phys.* **125**, 114501 (2006).
8. B. B. Laird, W. H. Thompson, *J. Chem. Phys.* **126**, 211104 (2007).
9. T. Fonseca, B. M. Ladanyi, *J. Phys. Chem.* **95**, 2116 (1991).
10. D. F. Underwood, D. A. Blank, *J. Phys. Chem. A* **109**, 3295 (2005).
11. D. S. Larsen, K. Ohta, Q. H. Xu, M. Cyrier, G. R. Fleming, *J. Chem. Phys.* **114**, 8008 (2001).
12. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Chem. Phys.* **122**, 134506 (2005).
13. D. T. Bowron, J. L. Finney, A. K. Soper, *J. Am. Chem. Soc.* **128**, 5119 (2006).
14. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Chem. Phys.* **125**, 194509 (2006).
15. I. B. Martini, E. R. Barthel, B. J. Schwartz, *Science* **293**, 462 (2001).
16. M. C. Cavanagh, R. E. Larsen, B. J. Schwartz, *J. Phys. Chem. A* **111**, 5144 (2007).
17. R. Catterall, J. Slater, M. C. R. Symons, *J. Chem. Phys.* **52**, 1003 (1970).
18. B. Bockrath, L. M. Dorfman, *J. Phys. Chem.* **77**, 1002 (1973).
19. F.-Y. Jou, G. R. Freeman, *Can. J. Chem.* **57**, 591 (1979).
20. A. E. Bragg, B. J. Schwartz, *J. Phys. Chem. B* **112**, 483 (2008).
21. A. E. Bragg, B. J. Schwartz, *J. Phys. Chem. A* **112**, 3530 (2008).
22. Methods are detailed in supporting material available on *Science* Online.
23. M. T. Lok, J. L. Dye, F. J. Tehan, *J. Phys. Chem.* **76**, 2975 (1972).
24. E. R. Barthel, B. J. Schwartz, *Chem. Phys. Lett.* **375**, 435 (2003).
25. O. Shoshana, J. L. Pérez Lustres, N. P. Ernsting, S. Ruhman, *Phys. Chem. Chem. Phys.* **8**, 2599 (2006).
26. V. Nagarajan, A. M. Brearley, T.-J. Kang, P. F. Barbara, *J. Chem. Phys.* **86**, 3183 (1987).
27. L. Reynolds, J. A. Gardecki, S. J. V. Frankland, M. L. Horng, M. Maroncelli, *J. Phys. Chem.* **100**, 10337 (1996).
28. A. C. Moskun, A. E. Jailaubekov, S. E. Bradforth, G. Tao, R. M. Stratt, *Science* **311**, 1907 (2006).
29. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Phys. Chem. A* **107**, 4773 (2003).
30. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *Phys. Rev. Lett.* **97**, 130403 (2006).
31. This research was funded by NSF under grant number CHE-0603766. The authors thank R. E. Larsen for useful discussions and a critical reading of the manuscript.

Mars' Paleomagnetic Field as the Result of a Single-Hemisphere Dynamo

Sabine Stanley,[1]* Linda Elkins-Tanton,[2] Maria T. Zuber,[2] E. Marc Parmentier[3]

Mars' crustal magnetic field was most likely generated by dynamo action in the planet's early history. Unexplained characteristics of the field include its strength, concentration in the southern hemisphere, and lack of correlation with any surface features except for the hemispheric crustal dichotomy. We used numerical dynamo modeling to demonstrate that the mechanisms proposed to explain crustal dichotomy formation can result in a single-hemisphere dynamo. This dynamo produces strong magnetic fields in only the southern hemisphere. This magnetic field morphology can explain why Mars' crustal magnetic field intensities are substantially stronger in the southern hemisphere without relying on any postdynamo mechanisms.

The Mars Global Surveyor mission showed that Mars possesses remanent crustal magnetic fields from a dynamo that was operational for a short time in Mars' early history (1). Remanent crustal magnetism is observed in early Noachian (>3.9 billion years old) crust in both

the northern and southern hemispheres, except for much of the Tharsis volcanic province and the large impact basins Hellas and Argyre in the southern hemisphere, and Isidis and Utopia in the northern hemisphere. There is a conspicuous difference in the magnetic field intensities in the two hemispheres: The northern hemisphere contains only weak magnetic fields, whereas the southern hemisphere contains both strong and weak fields (2).

The timing of the dynamo is constrained by the observations that the floors of the large impact basins formed during the Late Heavy Bombardment [~3.9 billion years ago (Ga)] are not

magnetized (1) and that the ancient Martian meteorite ALH84001 contains a remanent magnetic field dated earlier than 3.9 Ga (3). Most likely, the dynamo was active sometime between core formation (~4.5 Ga) and the Late Heavy Bombardment. The driving force for the dynamo, the intensity and morphology of the generated field, and the cause of the dynamo's demise are not well understood.

Another ancient Martian crustal feature is the hemispheric dichotomy. The northern and southern hemispheres have similar-aged crusts (4) but different topographies, thicknesses, and sediment covers (5). The northern hemisphere crust is low, thin, and covered with volcanic flows and sediments, whereas the southern hemisphere crust is high, thick, and largely devoid of sedimentary or volcanic resurfacing. Cratering evidence and the dichotomy's long wavelength suggest that the dichotomy is an ancient feature, directly related to crustal formation sometime between 4.5 and 3.9 Ga (6, 7).

Because the crustal magnetic field and the dichotomy are similar in age, it is possible that their formation processes are related. Several endogenic mechanisms could explain both dichotomy formation and a concurrent dynamo sometime between 4.5 and 3.9 Ga. A hemispheric-scale (degree-1) pattern of mantle circulation resulting from either mantle convection in the presence of

[1]Department of Physics, University of Toronto, Toronto, ON M5S1A7, Canada. [2]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Geological Sciences, Brown University, Providence, RI 02912, USA.

*To whom correspondence should be addressed. E-mail: stanley@physics.utoronto.ca

Mars' Paleomagnetic Field as the Result of a Single-Hemisphere Dynamo

reactions (Eqs. 3 to 5). Thus, although we may anticipate that the solvation dynamics measured involve substantial changes in the local solvation structure due to the solute size change, we expect that the energies associated with $(Na^+, e^-)^*$ formation are not chemically extreme and are representative of those associated with common solution-phase reactions.

This clear breakdown of LR implies that the solvent fluctuations coupled to the (Na^+, e^-) TCP are not Gaussian, and thus that the potential surfaces associated with these ET processes are highly nonparabolic. As a result, the Marcus theory of ET would poorly describe these ET processes. We anticipate that this could be an important consideration in many similar outer-sphere ET reactions, where substantial rearrangement of the local solvent structure could induce a similar LR breakdown. It is also important to note that the LR approximation is built on the idea that the same solvent-solute motions that underlie equilibrium fluctuations are also responsible for the nonequilibrium solvation dynamics (3). Although we observed a clear difference in the time dependence of two solvation pathways that reflects a breakdown of LR, observing an identical time dependence would not have guaranteed that the LR holds. This is because even when the specific molecular motions responsible for relaxing a nonequilibrium perturbation differ considerably from the solvent fluctuations active at equilibrium, LR may appear to be valid if the relevant nonequilibrium and equilibrium solvent

motions happen to occur on similar time scales; what we have termed a hidden breakdown of LR (29, 30). Overall, these findings demonstrate that an accurate assessment of solvation dynamics—and, by extension, our understanding of solution-phase chemical reactivity—must be considered directly at the molecular level in order to determine correctly how best to understand the solvent relaxation resulting from a given nonequilibrium perturbation.

References and Notes

1. R. A. Marcus, N. Sutin, *Biochim. Biophys. Acta* **811**, 265 (1985).
2. M. Maroncelli, *J. Mol. Liq.* **57**, 1 (1993).
3. D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford Univ. Press, New York, 1987).
4. R. M. Stratt, M. Maroncelli, *J. Phys. Chem.* **100**, 12981 (1996).
5. P. L. Geissler, D. Chandler, *J. Chem. Phys.* **113**, 9759 (2000).
6. D. Aherne, V. Tran, B. J. Schwartz, *J. Phys. Chem. B* **104**, 5382 (2000).
7. G. Tao, R. M. Stratt, *J. Chem. Phys.* **125**, 114501 (2006).
8. B. B. Laird, W. H. Thompson, *J. Chem. Phys.* **126**, 211104 (2007).
9. T. Fonseca, B. M. Ladanyi, *J. Phys. Chem.* **95**, 2116 (1991).
10. D. F. Underwood, D. A. Blank, *J. Phys. Chem. A* **109**, 3295 (2005).
11. D. S. Larsen, K. Ohta, Q. H. Xu, M. Cyrier, G. R. Fleming, *J. Chem. Phys.* **114**, 8008 (2001).
12. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Chem. Phys.* **122**, 134506 (2005).
13. D. T. Bowron, J. L. Finney, A. K. Soper, *J. Am. Chem. Soc.* **128**, 5119 (2006).
14. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Chem. Phys.* **125**, 194509 (2006).
15. I. B. Martini, E. R. Barthel, B. J. Schwartz, *Science* **293**, 462 (2001).
16. M. C. Cavanagh, R. E. Larsen, B. J. Schwartz, *J. Phys. Chem. A* **111**, 5144 (2007).
17. R. Catterall, J. Slater, M. C. R. Symons, *J. Chem. Phys.* **52**, 1003 (1970).
18. B. Bockrath, L. M. Dorfman, *J. Phys. Chem.* **77**, 1002 (1973).
19. F.-Y. Jou, G. R. Freeman, *Can. J. Chem.* **57**, 591 (1979).
20. A. E. Bragg, B. J. Schwartz, *J. Phys. Chem. B* **112**, 483 (2008).
21. A. E. Bragg, B. J. Schwartz, *J. Phys. Chem. A* **112**, 3530 (2008).
22. Methods are detailed in supporting material available on *Science* Online.
23. M. T. Lok, J. L. Dye, F. J. Tehan, *J. Phys. Chem.* **76**, 2975 (1972).
24. E. R. Barthel, B. J. Schwartz, *Chem. Phys. Lett.* **375**, 435 (2003).
25. O. Shoshana, J. L. Pérez Lustres, N. P. Ernsting, S. Ruhman, *Phys. Chem. Chem. Phys.* **8**, 2599 (2006).
26. V. Nagarajan, A. M. Brearley, T.-J. Kang, P. F. Barbara, *J. Chem. Phys.* **86**, 3183 (1987).
27. L. Reynolds, J. A. Gardecki, S. J. V. Frankland, M. L. Horng, M. Maroncelli, *J. Phys. Chem.* **100**, 10337 (1996).
28. A. C. Moskun, A. E. Jailaubekov, S. E. Bradforth, G. Tao, R. M. Stratt, *Science* **311**, 1907 (2006).
29. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *J. Phys. Chem. A* **107**, 4773 (2003).
30. M. J. Bedard-Hearn, R. E. Larsen, B. J. Schwartz, *Phys. Rev. Lett.* **97**, 130403 (2006).
31. This research was funded by NSF under grant number CHE-0603766. The authors thank R. E. Larsen for useful discussions and a critical reading of the manuscript.

Mars' Paleomagnetic Field as the Result of a Single-Hemisphere Dynamo

Sabine Stanley,[1]* Linda Elkins-Tanton,[2] Maria T. Zuber,[2] E. Marc Parmentier[3]

Mars' crustal magnetic field was most likely generated by dynamo action in the planet's early history. Unexplained characteristics of the field include its strength, concentration in the southern hemisphere, and lack of correlation with any surface features except for the hemispheric crustal dichotomy. We used numerical dynamo modeling to demonstrate that the mechanisms proposed to explain crustal dichotomy formation can result in a single-hemisphere dynamo. This dynamo produces strong magnetic fields in only the southern hemisphere. This magnetic field morphology can explain why Mars' crustal magnetic field intensities are substantially stronger in the southern hemisphere without relying on any postdynamo mechanisms.

The Mars Global Surveyor mission showed that Mars possesses remanent crustal magnetic fields from a dynamo that was operational for a short time in Mars' early history (1). Remanent crustal magnetism is observed in early Noachian (>3.9 billion years old) crust in both

[1]Department of Physics, University of Toronto, Toronto, ON M5S1A7, Canada. [2]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Geological Sciences, Brown University, Providence, RI 02912, USA.

*To whom correspondence should be addressed. E-mail: stanley@physics.utoronto.ca

the northern and southern hemispheres, except for much of the Tharsis volcanic province and the large impact basins Hellas and Argyre in the southern hemisphere, and Isidis and Utopia in the northern hemisphere. There is a conspicuous difference in the magnetic field intensities in the two hemispheres: The northern hemisphere contains only weak magnetic fields, whereas the southern hemisphere contains both strong and weak fields (2).

The timing of the dynamo is constrained by the observations that the floors of the large impact basins formed during the Late Heavy Bombardment [~3.9 billion years ago (Ga)] are not

magnetized (1) and that the ancient Martian meteorite ALH84001 contains a remanent magnetic field dated earlier than 3.9 Ga (3). Most likely, the dynamo was active sometime between core formation (~4.5 Ga) and the Late Heavy Bombardment. The driving force for the dynamo, the intensity and morphology of the generated field, and the cause of the dynamo's demise are not well understood.

Another ancient Martian crustal feature is the hemispheric dichotomy. The northern and southern hemispheres have similar-aged crusts (4) but different topographies, thicknesses, and sediment covers (5). The northern hemisphere crust is low, thin, and covered with volcanic flows and sediments, whereas the southern hemisphere crust is high, thick, and largely devoid of sedimentary or volcanic resurfacing. Cratering evidence and the dichotomy's long wavelength suggest that the dichotomy is an ancient feature, directly related to crustal formation sometime between 4.5 and 3.9 Ga (6, 7).

Because the crustal magnetic field and the dichotomy are similar in age, it is possible that their formation processes are related. Several endogenic mechanisms could explain both dichotomy formation and a concurrent dynamo sometime between 4.5 and 3.9 Ga. A hemispheric-scale (degree-1) pattern of mantle circulation resulting from either mantle convection in the presence of

radial viscosity variations (8, 9), early magma ocean crystallization resulting in overturn (10, 11), or superplumes resulting from destabilization of the mantle lower thermal boundary layer (12) provides degree-1 crustal structure along with

Table 1. Nondimensional parameter values in the dynamo model. Model values are given for the Prandtl number ($Pr = \upsilon/\kappa$, where υ is the kinematic viscosity and κ is the thermal diffusivity); a magnetic Prandtl number ($q_\kappa = \kappa/\eta$ where η is the magnetic diffusivity); the Ekman number [$E = \upsilon/(2\Omega r_o^2)$, where Ω is the angular rotation rate of the planet and r_o is the core radius]; and the modified Rayleigh number [$Ra = \alpha g_o h_T r_o^2/(2\Omega\eta)$, where α, g_o, and h_T are the thermal expansion coefficient, the gravitational acceleration at the CMB, and the prescribed superadiabatic temperature gradient at the inner core boundary, respectively]. Using representative Mars values of $\alpha = 10^{-5}$ K^{-1}, $g_o = 3.5$ m s^{-2}, $r_o = 1700$ km, $\Omega = 7.1 \times 10^{-5}$ s^{-1}, and $\eta = 2$ m s^{-2}, combined with the chosen value for the Rayleigh number, implies a superadiabatic temperature gradient at the inner core boundary of 5×10^{-8} K m^{-1}, corresponding to a superadiabatic temperature gradient at the CMB of $h_T r_{io}^2 = 6 \times 10^{-9}$ K m^{-1} (where r_{io} is the inner to outer core radius ratio).

Parameter	Value
Pr	1
q_κ	1
E	2×10^{-5}
Ra	18,000

sufficiently vigorous core convection to sustain a short-lived dynamo. Effects from a very large impact (13) or several large impacts (14) were also suggested early on but encountered difficulties (15, 16). The probability of several large impacts forming the northern lowlands is low, and there is no evidence of individual basins. A single very large impact is statistically possible. Presently, the dichotomy boundary is not circular or elliptical; however, a recent analysis of crustal thickness (17) demonstrated that its original shape was elliptical. Therefore, the dichotomy boundary and northern lowlands could be the result of a giant low-angle impact.

Although explanations for dichotomy formation with concurrent magnetic field generation appear feasible, a serious problem is the markedly different intensities of the crustal magnetic fields in the northern and southern hemispheres. If the dynamo produced an axial-dipole-dominated magnetic field, similar to the geomagnetic field, and the northern and southern crusts formed at similar times with similar magnetic mineral densities and magnetic layer thicknesses, then one would expect crustal fields of similar strength in both hemispheres. Efforts to explain the hemispheric magnetic intensity differences generally involve postdynamo mechanisms in the northern hemisphere, such as hydrothermal alteration (6) or demagnetization resulting from early large impacts (15).

Here, we show that dynamo generation can also explain the hemispheric magnetic intensity differences, thereby removing the requirement for a postdynamo solution. All endogenic mechanisms

involve hemispheric-scale mantle circulation, which will necessarily produce a degree-1 temperature anomaly in the mantle and hence at the core-mantle boundary (CMB). Numerical models have demonstrated that an exogenic (giant impact) mechanism could also produce a degree-1 temperature anomaly in the mantle and at the CMB (18). Because the CMB is the outer-bounding surface of the dynamo region, this temperature anomaly will result in a hemispheric heat flux variation on the outer boundary of the core. We therefore imposed a degree-1 variable heat flux pattern at the CMB in a dynamo simulation.

We used the numerical dynamo model of Kuang and Bloxham (19–21) with the parameter values given in Table 1. We imposed a heat flux across the CMB that was lower in the northern hemisphere than in the southern hemisphere (fig. S1). The heat flux variation on the CMB is relatively large, with the root mean square of the lateral variations three times the average superadiabatic heat flux. This may be reasonable for Mars when considering the relative temperatures of downwellings and upwellings at the CMB (11).

This spatially variable heat flux boundary condition produces a stable one-hemisphere dynamo. The radial component of the magnetic field is strongest and concentrated in the southern hemisphere with only weak fields in the northern hemisphere (Fig. 1A). Oscillatory hemispheric dynamos have been found in certain parameter regimes (22), producing a strong field in each hemisphere periodically. The dynamo in our simulation is different in that fields are actively gen-
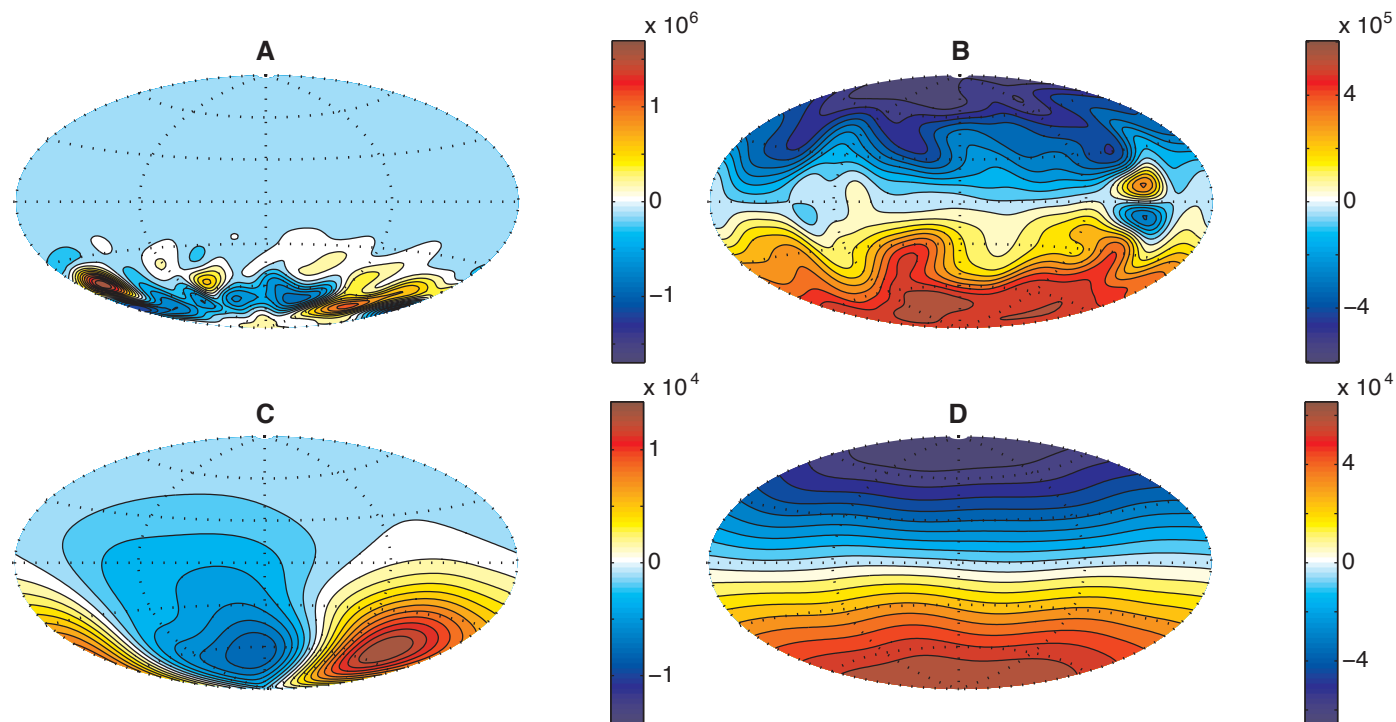
Fig. 1. Filled contours of the radial component of the magnetic field. For a model with a degree-1 heat flux outer-boundary condition, the field is plotted at the CMB in (**A**) and at the surface in (**C**). For a homogeneous heat flux outer-boundary condition, the field is plotted at the CMB in (**B**) and at the surface in (**D**). The CMB radius is 1700 km, the surface radius is 3400 km, and the units are nT.

erated in only a single hemisphere. The fields are smaller-scale and, although highly time-variable in morphology, remain strong in only the southern hemisphere (fig. S2). A dynamo model with a similar variable heat flux boundary condition as our model but with a different choice of velocity boundary condition does not produce a single-hemisphere dynamo (23), most likely because of the different force balances in the models [supporting online material (SOM) text].

The radial field at the CMB in our model was more intense than that in a model with the same parameter values but with a homogeneous heat flux boundary condition that produced an axial-dipolar dynamo (Fig. 1B). However, because the power in the field components fell off faster with distance for smaller scales, the magnetic field at the surface was slightly weaker in our variable heat flux model than in the homogeneous heat flux model but was the same order of magnitude (Fig. 1, C and D).

The hemispheric boundary condition in the model changed the equatorial symmetry of the superadiabatic temperature (Fig. 2), affecting the dominant force balance in the core. This had a substantial effect on the velocity fields of the core in our model (Fig. 3A). The zonal flows, mainly resulting from thermal winds, were equatorially antisymmetric and therefore did not adhere to the Taylor-Proudman constraint, which is expected to

hold when Coriolis and pressure forces dominate in the core (24). In addition, the meridional circulation pattern was concentrated in the southern hemisphere rather than filling the whole core, as in a model with homogeneous heat flux boundary conditions (Fig. 3B). The convection rolls were concentrated in the hemisphere with the colder CMB temperature and generated the strongest dynamo action in this region.

Our results suggest that the concentration of strong crustal fields in the southern hemisphere of Mars could result from a dynamo that produced a magnetic field concentrated in the southern hemisphere. In this scenario, no postdynamo process is required to remove a strong crustal field in the northern hemisphere. Although large basins are demagnetized in the northern hemisphere (as they are in the southern hemisphere), this mechanism can explain why none of the magnetized regions
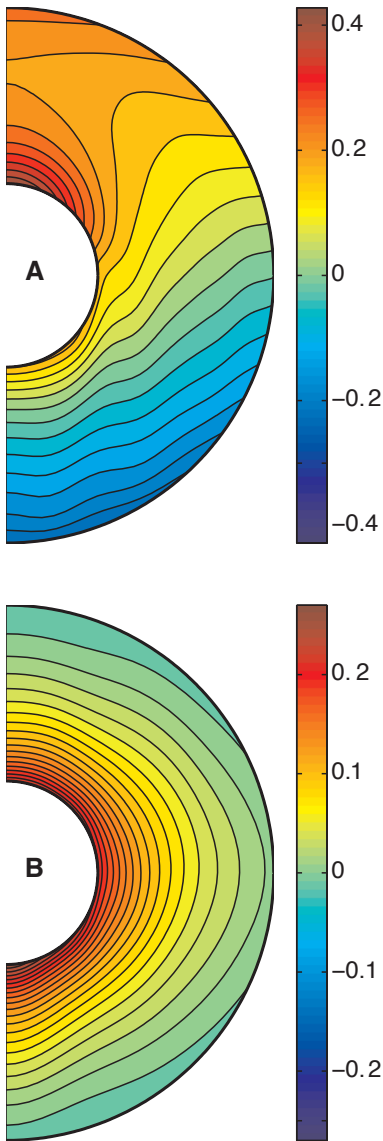
Fig. 2. Filled contours of the axisymmetric part of the nondimensional superadiabatic temperature. The profile for a model with a degree-1 heat flux outer-boundary condition is shown in (**A**) and for a model with a homogeneous heat flux outer-boundary condition in (**B**). The temperature plotted is with respect to the average CMB temperature, so positive values are hotter than the average CMB temperature and negative values are colder than the average CMB temperature. To dimensionalize in Kelvin, nondimensional values should be multiplied by 0.085.

Fig. 3. Axisymmetric component of the non-dimensional velocity field. Filled contours of the zonal velocity are shown on the left, and stream lines of the meridional circulation are shown on the right for a model with a degree-1 heat flux outer-boundary condition (**A**) and a model with a homogeneous heat flux outer-boundary condition (**B**). For the zonal velocity, red indicates prograde zonal flow, and blue, retrograde zonal flow. For the meridional circulation, red indicates prograde circulation, and blue, retrograde circulation. The color bars apply to the zonal velocity figures only, and the units are m s^{-1}. The intensity of the meridional circulation is prescribed by the spacing of the stream lines.

in the northern hemisphere are as strong as the regions in the south.

In addition to explaining the occurrence of a strong field in only one hemisphere, our model is also able to explain the conflicts between various Mars paleomagnetic studies and rotational stability studies. The inferred paleomagnetic pole positions vary in location depending on the individual crustal anomaly used (25, 26). Some of the paleopoles are also located in equatorial regions near the Tharsis bulge and hence are far from the current geographic poles. This has been interpreted as evidence for a large true polar wander event that relocated Tharsis from polar to equatorial regions in early Mars history (27). However, large Tharsis-driven true polar wander is in conflict with rotational stability studies (28, 29) that demonstrate that the present-day gravitational figure of Mars favors a small Tharsis-driven true-polar wander scenario.

An assumption made in the paleomagnetic studies is that the dynamo-generated magnetic field was axial-dipolar dominated. This assumption implies that the magnetic pole coincided with the rotation pole and is used extensively in Earth paleomagnetic studies. Our models would dictate that the Mars dynamo-generated field was not axial-dipolar dominated and hence that the magnetic poles would not coincide with the rotation poles, rendering paleopole interpretations useless. In addition, because our dynamo-generated fields are multipolar, individual crustal magnetic fields at different locations can point to different paleomagnetic poles, thereby explaining the discrepancies in the different paleomagnetic studies.

A single-hemisphere dynamo also has implications for evolution of the martian atmosphere. A strong dynamo-generated magnetic field can more easily explain the intense crustal magnetism. However, efficient atmospheric erosion, necessary to explain the loss of Mars' early thick atmosphere, favors a weak internal magnetic field (30). Our single-hemisphere dynamo may provide an elegant solution to this problem because the northern hemisphere would be prone to atmospheric removal early in solar system history when the young Sun was more active (31), but the southern hemisphere could still possess a strong magnetic field in which the crustal rocks could magnetize.

References and Notes

1. M. H. Acuna et al., Science 284, 790 (1999).
2. B. Langlais, M. E. Purucker, M. Mandea, J. Geophys. Res. 109, E02008 (2004).
3. B. P. Weiss et al., Earth Planet. Sci. Lett. 201, 449 (2002).
4. L. A. Edgar, H. V. Frey, Geophys. Res. Lett. 35, L02201 (2008).
5. M. T. Zuber et al., Science 287, 1788 (2000).
6. S. C. Solomon et al., Science 307, 1214 (2005) and references therein.
7. H. V. Frey, J. Geophys. Res. 111, E08S91 (2006).
8. S. Zhong, M. T. Zuber, Earth Planet. Sci. Lett. 189, 75 (2001).
9. J. H. Roberts, S. Zhong, J. Geophys. Res. 111, E06013 (2006).
10. L. T. Elkins-Tanton, E. M. Parmentier, P. C. Hess, Meteorit. Planet. Sci. 38, 1753 (2003).
11. L. T. Elkins-Tanton, S. E. Zaranek, E. M. Parmentier, P. C. Hess, Earth Planet. Sci. Lett. 236, 1 (2005).
12. Y. Ke, V. S. Solomatov, J. Geophys. Res. 111, E10001 (2006).
13. D. E. Wilhelms, S. W. Squyres, Nature 309, 138 (1984).
14. H. Frey, R. A. Shultz, Geophys. Res. Lett. 15, 229 (1988).
15. F. Nimmo, M. S. Gilmore, J. Geophys. Res. 106, 12315 (2001).
16. T. R. Watters, P. J. McGovern, R. P. Irwin III, Annu. Rev. Earth Planet. Sci. 35, 621 (2007).
17. J. C. Andrews-Hanna, M. T. Zuber, W. B. Banerdt, Nature 453, 1212 (2008).
18. W. A. Watters, M. T. Zuber, B. H. Hager, J. Geophys. Res., in press.
19. W. Kuang, J. Bloxham, Nature 389, 371 (1997).
20. W. Kuang, J. Bloxham, J. Comput. Phys. 153, 51 (1999).
21. Additional information on the numerical methods is available as supporting material on Science Online.
22. E. Grote, F. H. Busse, Phys. Rev. E 62, 4457 (2000).
23. G. A. Glatzmaier, R. S. Coe, L. Hongre, P. H. Roberts, Nature 401, 885 (1999).
24. M. Kono, P. H. Roberts, Rev. Geophys. 40, 1013 (2002).
25. Y. Quesnel, B. Langlais, C. Sotin, Planet. Space Sci. 55, 258 (2007), and references therein.
26. B. Langlais, M. Purucker, Planet. Space Sci. 55, 270 (2007); and references therein.
27. K. F. Sprenke, L. L. Baker, A. F. Williams, Icarus 174, 486 (2005); and references therein.
28. A. Daradich et al., Icarus 194, 463 (2008).
29. J. T. Perron, J. X. Mitrovica, M. Manga, I. Matsuyama, M. A. Richards, Nature 447, 840 (2007).
30. V. Dehant et al., Space Sci. Rev. 129, 279 (2007).
31. Y. N. Kulikov et al., Space Sci. Rev. 129, 207 (2007).
32. S.S. is partially funded by the National Science and Engineering Research Council (NSERC) of Canada. The numerical simulations in this study were performed on supercomputing resources partially funded by the Canadian Foundation for Innovation (CFI) and the Ontario Research Fund (ORF).

The Structure and Dynamics of Mid-Ocean Ridge Hydrothermal Systems

D. Coumou,* T. Driesner, C. A. Heinrich

Sub-seafloor hydrothermal convection at mid-ocean ridges transfers 25% of the Earth's heat flux and can form massive sulfide ore deposits. Their three-dimensional (3D) structure and transient dynamics are uncertain. Using 3D numerical simulations, we demonstrated that convection cells self-organize into pipelike upflow zones surrounded by narrow zones of focused and relatively warm downflow. This configuration ensures optimal heat transfer and efficient metal leaching for ore-deposit formation. Simulated fluid-residence times are as short as 3 years. The concentric flow geometry results from nonlinearities in fluid properties, and this may influence the behavior of other fluid-flow systems in Earth's crust.

Hydrothermal convection at mid-ocean ridge spreading centers transports a major part of Earth's total heat flux, substantially affects the chemistry of crust and overlying ocean, and provides nutrients for chemosynthetic life on and beneath the sea floor. Mass, heat, and associated chemical fluxes from the crust to the ocean at mid-ocean ridge spreading centers are large (1, 2). Fundamental features of this flow, such as the location of seawater recharge and the relative importance of off- versus along-axis convection, are still uncertain. Recent studies of active (3) and ancient (4) systems show that discharge can be highly focused in pipelike regions, possibly continuing to the base of the hydrothermal system (3). Recharge is often thought to occur over extensive areas (5, 6), with off-axis faults guiding fluid pathways toward the base of the hydrothermal system. A common alternative view is that of fluid circulation being restricted to a high-permeability along-axis zone (7, 8). Micro-earthquake data indicate that recharge can be focused close to the spreading center in some systems (9).

Recent two-dimensional (2D) numerical studies that included accurate thermodynamic properties of water have shown that the nonlinear dependence of fluid properties on pressure and temperature is a first-order control, determining the self-organization of convection cells (10–12). Quantitative 3D numerical models have been applied to low-permeability (13) or sedimented systems (14) and to a configuration with an along-axis high-permeability fracture (15, 16) but not to the more highly permeable basaltic systems, which represent the greater and most active part of mid-ocean ridge spreading centers. Here, we describe a 3D model that represents the hydrothermal system without geological complexity so as to identify the first-order physical factors controlling the behavior of mid-ocean ridge convection cells.

The governing equations are an appropriate version of Darcy's law (17), conservation of mass and energy in an incompressible porous medium (12) and an accurate equation of state for pure water (18). Using pure water substantially reduces the computational complexity because pure water above the critical pressure (21.1 MPa) is always a single-phase fluid with properties closely resembling those of seawater. Two-dimensional across-axis simulations including the full-phase

Department of Earth Sciences, Eidgenössische Technische Hochschule–Zurich, Clausiusstrasse 25, Zurich 9082, Switzerland.

*To whom correspondence should be addressed. E-mail: coumou@pik-potsdam.de

The Structure and Dynamics of Mid-Ocean Ridge Hydrothermal Systems

D. Coumou, *et al.*
Science **321**, 1825 (2008);
DOI: 10.1126/science.1159582

in the northern hemisphere are as strong as the regions in the south.

In addition to explaining the occurrence of a strong field in only one hemisphere, our model is also able to explain the conflicts between various Mars paleomagnetic studies and rotational stability studies. The inferred paleomagnetic pole positions vary in location depending on the individual crustal anomaly used (25, 26). Some of the paleopoles are also located in equatorial regions near the Tharsis bulge and hence are far from the current geographic poles. This has been interpreted as evidence for a large true polar wander event that relocated Tharsis from polar to equatorial regions in early Mars history (27). However, large Tharsis-driven true polar wander is in conflict with rotational stability studies (28, 29) that demonstrate that the present-day gravitational figure of Mars favors a small Tharsis-driven true-polar wander scenario.

An assumption made in the paleomagnetic studies is that the dynamo-generated magnetic field was axial-dipolar dominated. This assumption implies that the magnetic pole coincided with the rotation pole and is used extensively in Earth paleomagnetic studies. Our models would dictate that the Mars dynamo-generated field was not axial-dipolar dominated and hence that the magnetic poles would not coincide with the rotation poles, rendering paleopole interpretations useless. In addition, because our dynamo-generated fields are multipolar, individual crustal magnetic fields at different locations can point to different paleomagnetic poles, thereby explaining the discrepancies in the different paleomagnetic studies.

A single-hemisphere dynamo also has implications for evolution of the martian atmosphere. A strong dynamo-generated magnetic field can more easily explain the intense crustal magnetism. However, efficient atmospheric erosion, necessary to explain the loss of Mars' early thick atmosphere, favors a weak internal magnetic field (30). Our single-hemisphere dynamo may provide an elegant solution to this problem because the northern hemisphere would be prone to atmospheric removal early in solar system history when the young Sun was more active (31), but the southern hemisphere could still possess a strong magnetic field in which the crustal rocks could magnetize.

References and Notes

1. M. H. Acuna et al., Science 284, 790 (1999).
2. B. Langlais, M. E. Purucker, M. Mandea, J. Geophys. Res. 109, E02008 (2004).
3. B. P. Weiss et al., Earth Planet. Sci. Lett. 201, 449 (2002).
4. L. A. Edgar, H. V. Frey, Geophys. Res. Lett. 35, L02201 (2008).
5. M. T. Zuber et al., Science 287, 1788 (2000).
6. S. C. Solomon et al., Science 307, 1214 (2005) and references therein.
7. H. V. Frey, J. Geophys. Res. 111, E08S91 (2006).
8. S. Zhong, M. T. Zuber, Earth Planet. Sci. Lett. 189, 75 (2001).
9. J. H. Roberts, S. Zhong, J. Geophys. Res. 111, E06013 (2006).
10. L. T. Elkins-Tanton, E. M. Parmentier, P. C. Hess, Meteorit. Planet. Sci. 38, 1753 (2003).
11. L. T. Elkins-Tanton, S. E. Zaranek, E. M. Parmentier, P. C. Hess, Earth Planet. Sci. Lett. 236, 1 (2005).
12. Y. Ke, V. S. Solomatov, J. Geophys. Res. 111, E10001 (2006).
13. D. E. Wilhelms, S. W. Squyres, Nature 309, 138 (1984).
14. H. Frey, R. A. Shultz, Geophys. Res. Lett. 15, 229 (1988).
15. F. Nimmo, M. S. Gilmore, J. Geophys. Res. 106, 12315 (2001).
16. T. R. Watters, P. J. McGovern, R. P. Irwin III, Annu. Rev. Earth Planet. Sci. 35, 621 (2007).
17. J. C. Andrews-Hanna, M. T. Zuber, W. B. Banerdt, Nature 453, 1212 (2008).
18. W. A. Watters, M. T. Zuber, B. H. Hager, J. Geophys. Res., in press.
19. W. Kuang, J. Bloxham, Nature 389, 371 (1997).
20. W. Kuang, J. Bloxham, J. Comput. Phys. 153, 51 (1999).
21. Additional information on the numerical methods is available as supporting material on Science Online.
22. E. Grote, F. H. Busse, Phys. Rev. E 62, 4457 (2000).
23. G. A. Glatzmaier, R. S. Coe, L. Hongre, P. H. Roberts, Nature 401, 885 (1999).
24. M. Kono, P. H. Roberts, Rev. Geophys. 40, 1013 (2002).
25. Y. Quesnel, B. Langlais, C. Sotin, Planet. Space Sci. 55, 258 (2007), and references therein.
26. B. Langlais, M. Purucker, Planet. Space Sci. 55, 270 (2007); and references therein.
27. K. F. Sprenke, L. L. Baker, A. F. Williams, Icarus 174, 486 (2005); and references therein.
28. A. Daradich et al., Icarus 194, 463 (2008).
29. J. T. Perron, J. X. Mitrovica, M. Manga, I. Matsuyama, M. A. Richards, Nature 447, 840 (2007).
30. V. Dehant et al., Space Sci. Rev. 129, 279 (2007).
31. Y. N. Kulikov et al., Space Sci. Rev. 129, 207 (2007).
32. S.S. is partially funded by the National Science and Engineering Research Council (NSERC) of Canada. The numerical simulations in this study were performed on supercomputing resources partially funded by the Canadian Foundation for Innovation (CFI) and the Ontario Research Fund (ORF).

The Structure and Dynamics of Mid-Ocean Ridge Hydrothermal Systems

D. Coumou,* T. Driesner, C. A. Heinrich

Sub-seafloor hydrothermal convection at mid-ocean ridges transfers 25% of the Earth's heat flux and can form massive sulfide ore deposits. Their three-dimensional (3D) structure and transient dynamics are uncertain. Using 3D numerical simulations, we demonstrated that convection cells self-organize into pipelike upflow zones surrounded by narrow zones of focused and relatively warm downflow. This configuration ensures optimal heat transfer and efficient metal leaching for ore-deposit formation. Simulated fluid-residence times are as short as 3 years. The concentric flow geometry results from nonlinearities in fluid properties, and this may influence the behavior of other fluid-flow systems in Earth's crust.

Hydrothermal convection at mid-ocean ridge spreading centers transports a major part of Earth's total heat flux, substantially affects the chemistry of crust and overlying ocean, and provides nutrients for chemosynthetic life on and beneath the sea floor. Mass, heat, and associated chemical fluxes from the crust to the ocean at mid-ocean ridge spreading centers are large (1, 2). Fundamental features of this flow, such as the location of seawater recharge and the relative importance of off- versus along-axis convection, are still uncertain. Recent studies of active (3) and ancient (4) systems show that discharge can be highly focused in pipelike regions, possibly continuing to the base of the hydrothermal system (3). Recharge is often thought to occur over extensive areas (5, 6), with off-axis faults guiding fluid pathways toward the base of the hydrothermal system. A common alternative view is that of fluid circulation being restricted to a high-permeability along-axis zone (7, 8). Micro-earthquake data indicate that recharge can be focused close to the spreading center in some systems (9).

Recent two-dimensional (2D) numerical studies that included accurate thermodynamic properties of water have shown that the nonlinear dependence of fluid properties on pressure and temperature is a first-order control, determining the self-organization of convection cells (10–12). Quantitative 3D numerical models have been applied to low-permeability (13) or sedimented systems (14) and to a configuration with an along-axis high-permeability fracture (15, 16) but not to the more highly permeable basaltic systems, which represent the greater and most active part of mid-ocean ridge spreading centers. Here, we describe a 3D model that represents the hydrothermal system without geological complexity so as to identify the first-order physical factors controlling the behavior of mid-ocean ridge convection cells.

The governing equations are an appropriate version of Darcy's law (17), conservation of mass and energy in an incompressible porous medium (12) and an accurate equation of state for pure water (18). Using pure water substantially reduces the computational complexity because pure water above the critical pressure (21.1 MPa) is always a single-phase fluid with properties closely resembling those of seawater. Two-dimensional across-axis simulations including the full-phase

Department of Earth Sciences, Eidgenössische Technische Hochschule–Zurich, Clausiusstrasse 25, Zurich 9082, Switzerland.

*To whom correspondence should be addressed. E-mail: coumou@pik-potsdam.de

relations of seawater have shown that the overall narrow shape of the upwelling plume remains similar in spite of phase separation (*19*), hence the simplification employed here is justified. The equations are solved within the 3D box shown in Fig. 1, which has been discretized by a total of ~2.5 million tetrahedral elements. The total amount of energy added to the system can be estimated from the spreading rate of the oceanic plates. Magma crystallization and cooling to ambient temperatures release energy up to ~120 MW/km of the spreading axis (*2*). Direct measurements from individual vent fields or ridge segments have given substantially larger values, indicating that magma supply is episodic and local (*20–22*). In the model, we set the heat flux at an intermediate value of 350 MW/km, which is distributed along the bottom in a Gaussian profile that mimics an elongated magma chamber with an across-axial

width of roughly 1 km (Fig. 1) (*23*). We set a bulk permeability of $k = 5.0 \times 10^{-14}$ m^2. At these specific values of permeability and heat flux, the temperature near the bottom of the model establishes itself to be close to the expected magmatic temperatures of ~1200°C (*24*). At lower permeability, the system cannot remove heat fast enough, and the system heats up unrealistically. At higher permeabilities, heat is mined faster, suppressing bottom boundary temperatures to values much smaller than magmatic temperatures. The top boundary represents the sea floor and is kept at a pressure of 25 MPa, representing an ocean depth of roughly 2.5 km. To allow hot fluids to vent through the top, we used a mixed thermal boundary condition (*11*, *12*): In elements experiencing upflow, the thermal gradient is set to zero to represent unobstructed outflow to the ocean, whereas elements experiencing downflow take in 10°C water.

In the simulations, after a short initialization period, convection developed into pipelike upflow regions, spaced at regular distances of roughly 500 m (Fig. 2A). Although thermal instabilities periodically form and rise close to the axial center, the locations of the main upflow regions remain relatively fixed. Within these pipes, fluids of ~400°C move upward vigorously and vent at the sea floor, forming near-circular discharge areas. The surface area of these fields is ~2 × 10^4 m^2, which is in good agreement with the range observed for natural black smoker fields of 3 × 10^3 m^2 to 1 × 10^5 m^2 (*25*). Most of the downflow happens in concentric tube-shaped regions directly surrounding the upflow plumes (Fig. 2, A to C). Fluid temperatures in these regions ranged from 100° to 300°C at already shallow depth. At a radius of less than twice that of the upflow zone, the downward mass flux reached a value five times as large as the average downward mass flux and about half the maximum upward flux. At a depth of 500 m, the integrated upward mass flux through the whole model is ~800 kg/s, which implies ~200 kg/s per kilometer of ridge segment and ~100 kg/s per vent field. Roughly two-thirds of this mass is provided by downflow in the warm near-axial regions. Elevated temperatures of the downflowing fluids reaching 300°C imply that substantial heat is recirculated; roughly one-quarter of the heat traveling upward is lost conductively to the adjacent downflow zones, in which it is recirculated back to the bottom. This recirculation process therefore controls the thermal structure of the downflow zone, heating it to average temperatures of ~200°C. Relatively warm downflow has been observed in 2D models (*26*), in which downflow has to take place in between

Fig. 1. 3D mesh consisting of 2.5-m tetrahedron elements. Resolution is refined toward the axial center, with the finest resolution between the dashed lines, and colors indicate computational domains assigned to separate processors (*23*). At the bottom boundary, the bell-shaped form of the heat flux profile is plotted, ranging from 10 W/m^2 (blue) to 300 W/m^2 (red).

Fig. 2. Thermal and fluid-flow structure of the simulation after ~100 years. (**A**) The 3D contours of the 300°C (transparent) and 380°C (yellow) isotherms. The horizontal cross-section plots vertical mass fluxes [scale as plotted in (B)]. (**B**) Mass fluxes in kilograms per square meter per second through horizontal cross-sections at 100- and 500-m depth. Thin black lines plot the 100°C and 300°C isotherms, and the white line separates the upflowing from downflowing regions (*x* axis, meters; *y* axis, meters). (**C**) Cross-section through a thermal plume plotting 100°C (blue), 300°C (brown), and 380°C (yellow) isotherms as well as mass fluxes (arrows). High-hydraulic-conductivity downflow zones between the 100° and 300°C isotherms are clearly visible.

hot upflow zones, but is not naturally expected in 3D. Darcy fluxes in the system are $\sim 5.0 \times 10^{-6}$ m/s in the upflow zone and $\sim 0.75 \times 10^{-6}$ m/s in the surrounding near-axial downflow zone. By converting these values to actual pore velocities, a residence time of only ~ 3 years is calculated for a fluid particle entering the system through the proximal recharge zone, traveling to the base of the hydrothermal cell, and flowing out again through the black smoker. This short travel time is in agreement with recent estimates made with natural radionuclide tracers (27, 28).

To understand why this system evolves this way, we analyzed the steady-state pressure gradient ($\partial p/\partial z$) in a simplified geometry, taking account of the nonlinearity of the fluid properties. In the near-axial region, the vertical pressure gradient must have a value between that of a cold and hot hydrostatic, so that cold fluids can be brought downward and hot fluids can be brought upward. As a first approximation, we assumed that recharge is limited to a region directly surrounding a pipe-shaped upflow zone (fig. S1). The mass balance between up- and downflowing water is

$$A_u \rho_u \frac{k_u}{\mu_u}\left[\frac{\partial p}{\partial z} - \rho_u g\right] =$$
$$A_d \rho_d \frac{k_d}{\mu_d}\left[\rho_d g - \frac{\partial p}{\partial z}\right] \quad (1)$$

Here, A_u is the cross-sectional area of the upflow zone and A_d is the cross-sectional area of the downflow zone. Subscripts u and d indicate properties of the upwelling and the downwelling fluid, respectively; ρ is the water density; μ is the water viscosity; and g is the acceleration due to gravity (fig. S1). Equation 1 assumes that horizontal pressure gradients between up- and downflow areas can be neglected. This assumption is allowed in the near-axial region because there the fluid flow is dominantly vertical. The vertical pressure gradient can be expressed in terms of the density ρ_n of a fictitious neutrally buoyant fluid, according to $\partial p/\partial z = \rho_n g$. Equation 1 can now be written as

$$\frac{\rho_n - \rho_u}{\rho_d - \rho_u} = \frac{1}{1 + \gamma R} \quad (2)$$

Here, the geometric constant γ represents the ratio $A_u k_u / A_d k_d$ and R is the ratio of fluid prop-

erties $\mu_d \rho_u / \mu_u \rho_d$. Physically, the product γR can be seen as the ratio of the resistance against flow in the downflow region over that in the upflow region and, depending on pressure and temperature, can have values ranging from 0.1 to 10 γ. From Eq. 2, we can derive an expression for the fluxibility F, defined as the ability of a system to transport energy by buoyancy-driven convection

$$F = \frac{\rho_u (h_u - h_d)(\rho_d - \rho_u)}{\mu_u} \frac{1}{1 + \gamma R} \quad (3)$$

where h is the specific enthalpy of the fluid. Equation 3 is an extension of the ordinary definition of fluxibility (29) because it allows resistance in the downwelling limb as well as the upwelling limb. In the limit $\gamma \to 0$, in which the resistance of the downwelling limb is negligible as compared with that of the upwelling limb, Eq. 3 reduces to the original definition of fluxibility (29, 30).

Heating water from 10° to 200°C decreases μ by approximately one order of magnitude, whereas ρ changes by only about 10% (Fig. 3A). As a consequence, 200°C fluids can be brought downward highly efficiently: The greatly reduced viscosity makes them mobile, whereas their only weakly decreased density maintains a substantial downward buoyancy force as compared with the upflow zone. Heating the fluid to even higher temperatures would cause the density to decrease substantially and hence lower the downward buoyancy force. Downward mass transport can therefore be optimized at temperatures of around 200°C.

In order to analyze our numerical results, we took γ as ~ 1 because k is homogeneous ($k_d = k_u$), and the area of upflow is roughly the same as the area of increased downflow (Fig. 2B). Figure 3B plots F against temperatures of the up- and downflowing fluid using $\gamma = 1$. It shows that energy transport can be maximized when upflowing water reaches temperatures of $\sim 400°C$ and downflowing water is $\sim 200°C$, exactly like the thermal structure observed in the numerical simulations. The thermal structure of the numerical simulations featured enhanced downflow for fluids in the temperature range from 100° to 300°C, which fits with the peak of maximum fluxibility in Fig. 3B. At these conditions, a neutrally buoyant fluid has a density of ~ 600 kg/m^3 (Fig. 3C), implying that $\partial p/\partial z$ is ~ 6000 Pa/m, which is in agreement with simulation results. If upflowing fluids have either

a lower or higher temperature than $\sim 400°C$, then their hydraulic resistance will increase, hence the pressure gradient has to increase to drive such fluids upwards, resulting in larger neutral buoyancies (Fig. 3C). Similarly, the resistance of the downflow zone increases if its temperature is either larger or smaller than $\sim 200°C$. Cooling a $\sim 200°C$ fluid will increase its viscosity more rapidly than its density, whereas heating a $\sim 200°C$ fluid will decrease its density faster than its viscosity (see Fig. 3A). Both effects result in a larger hydraulic resistance of the fluid. This analysis shows that the convection cells operate in a state of least hydraulic resistance in order to maximize the overall efficiency of energy transport. Thus, surprisingly, substantial recirculation of heat unexpectedly enhances mass fluxes through the system and thereby increases the overall heat transport of the cell. Previous studies indicated that black smoker systems could well be operating in a state of maximum energy transport (30–33), explaining why black smoker vent fields are often close to but never higher than $\sim 400°C$ (11).

Convection cells evolve toward and remain in this state of least resistance by counteracting feedback mechanisms that self-regulate the flow in convection cells. For instance, increasing resistance by cooling the downflow zone to temperatures lower than $\sim 200°C$ slows down the fluid flow through the system. This reduced flow increases conductive heat losses from the upflow area, heating up the downflow area again to temperatures of $\sim 200°C$. Furthermore, the upflow temperature cannot become hotter than $\sim 400°C$ (11). Therefore, if convection stagnates because of external forces, the boundary layer between the magma chamber and hydrothermal cell will broaden, but upflow temperatures will stay at $\sim 400°C$.

The regular spacing between black smoker fields has previously been explained with a slot-convection model, in which convection dominantly takes place in a high-permeability axial plane and therefore could be considered a 2D system (12). Our simulations show that this restriction is not required and that, at high heat-flux conditions, convection naturally forms individual black smoker fields at regular spacing because of the nonlinear temperature and pressure dependence of the fluid's properties.

The Main Endeavour Field (MEF) on the Juan de Fuca ridge appears to be a natural exam-

Fig. 3. (**A**) Density (solid) and viscosity (dashed) as a function of temperature at 30 MPa. (**B**) F plotted against upwelling and downwelling temperatures (Eq. 3). Solid black lines follow contours at downwelling temperature (T_d) = 100°C, T_d = 200°C, T_d = 300°C, and T_u = 400°C. Blue, 0 < F < 1e15; Green, 1e15 < F < 2e15; Yellow, 2e15 < F < 3e15; and Orange, 3e15 < F < 4e15. (**C**) Neutral density (in kilograms per cubic meter) plotted against upwelling and downwelling temperatures.

ple closely matching our simulation results. There, regularly spaced vent sites, with black smokers venting fluids close to 400°C, have been active for several years (*34*, *35*). Magnetic anomaly data show that upflow zones are narrow pipelike structures that reach to a depth of at least a few hundreds of meters and possibly to the base of the hydrothermal system with a regular spacing comparable with that in our simulations (*3*). Though natural systems are geologically much more complex than the numerical model described here, we suggest that the MEF is likely to operate in a state of maximum energy transport, with recharge occurring close to the vent sites. The optimal site for future in situ tracer injection experiments would therefore be about twice the radius of a black smoker vent field. The relatively short residence times resulting from our calculations also indicate that such an experiment can lead to a successful tracer test within a realistic time frame.

Massive sulfide ore deposits form when convecting seawater leaches metals from new basaltic crust and reprecipitates them as sulfides at the outflow points of active black smokers. The solubility of iron, zinc, and copper increases almost exponentially with temperature (*36*). Our simulations indicate that the average temperature of fluid-rock interaction is much higher than would be expected from dispersed seawater infiltration across the spreading axis (*37*). As a result, solubility-limited leaching is expected to be much more effective by including large parts of the hot downflow path and not being restricted to the basal reaction zone of the system. With an average copper content in mid-ocean ridge basalt of 25 parts per million, a typical small seafloor sulfide deposit of 0.2 million metric tons containing 3% copper (*38*) can be formed by leaching only the

upflow zone and its immediately surrounding hot downflow zone. The copper solubility in this region varies from $\sim 10^{-7}$ mol/kg at 200°C to $\sim 10^{-4}$ mol/kg at 350°C, ensuring that the metals are quickly leached and a deposit can form within a period of 100 to 1000 years.

References and Notes
1. C. Stein, S. Stein, *J. Geophys. Res.* **99**, 3081 (1994).
2. A. Fisher, in *Energy and Mass Transfer in Marine Hydrothermal Systems*, P. E. Halbach, V. Tunnicliffe, J. R. Hein, Eds. (Dahlem Univ. Press, Berlin, 2003), vol. 89, pp. 29–52.
3. M. Tivey, H. Johnson, *Geology* **30**, 979 (2002).
4. L. Coogan et al., *Am. J. Sci.* **306**, 389 (2006).
5. H. Johnson, K. Becker, R. Von Herzen, *Geophys. Res. Lett.* **20**, 1875 (1993).
6. S. Kelley, J. Baross, J. Delaney, *Annu. Rev. Earth Planet. Sci.* **30**, 385 (2002).
7. P. Nehlig, T. Juteau, *Mar. Geol.* **84**, 209 (1988).
8. W. Wilcock, A. Fisher, *Geophys. Monogr.* **144**, 51 (2004).
9. M. Tolstoy, F. Waldhause, D. Bohnenstiehl, R. Weekly, W. Kim, *Nature* **451**, 181 (2008).
10. S. Ingebritsen, D. O. Hayba, *Geophys. Res. Lett.* **21**, 2199 (1994).
11. T. Jupp, A. Schultz, *Nature* **403**, 880 (2000).
12. D. Coumou, T. Driesner, S. Geiger, C. Heinrich, S. Matthai, *Earth Planet. Sci. Lett.* **245**, 218 (2006).
13. B. Travis, D. Janecky, N. Rosenberg, *Geophys. Res. Lett.* **18**, 1441 (1991).
14. M. Rabinowicz, J. Boulegue, P. Genthon, *J. Geophys. Res.* **103**, 24045 (1998).
15. M. Rabinowicz, J. Sempéré, P. Genthon, *J. Geophys. Res.* **104**, 29275 (1999).
16. F. Fontaine, M. Rabinowicz, J. Boulegue, *Earth Planet. Sci. Lett.* **184**, 407 (2001).
17. S. Ingebritsen, W. Sanford, C. Neuzil, *Groundwater in Geologic Processes* (Cambridge Univ. Press, Cambridge, ed. 2, 2006).
18. G. Haar, Kell, *NBS/NRC Steam Tables* (Hemisphere Publishing, New York, 1984).
19. D. Coumou, thesis, Eidgenössische Technische Hochschule–Zurich, Switzerland (2008).
20. U. Ginster, M. Mottl, R. Von Herzen, *J. Geophys. Res.* **99**, 4937 (1994).
21. E. Baker, T. Urabe, *J. Geophys. Res.* **101**, 8685 (1996).
22. P. Ramondenc, L. Germanovich, K. Von Damm, R. Lowell, *Earth Planet. Sci. Lett.* **245**, 487 (2006).
23. Materials and methods are available as supporting material on *Science* Online.
24. J. Sinton, R. Detrick, *J. Geophys. Res.* **97**, 197 (1992).
25. R. Lowell, A. Rona, R. Von Herzen, *J. Geophys. Res.* **100**, 327 (1995).
26. F. Fontaine, W. Wilcock, *Geochem. Geophys. Geosyst.* **8**, Q07010 (2007).
27. D. Kadko, D. Butterfield, *Geochim. Cosmochim. Acta* **62**, 1521 (1998).
28. D. Kadko, G. K., D. Butterfield. *Geochim. Cosmochim. Acta* **71**, 6019 (2007).
29. C. Lister, *Geophys. J. Int.* **120**, 45 (1995).
30. T. Jupp, A. Schultz, *J. Geophys. Res.* **109**, 10.1029/2003JB002697 (2004).
31. W. Malkus, *Proc. R. Soc. London. Ser. A* **225**, 196 (1954).
32. F. Busse, D. Joseph, *J. Fluid Mech.* **54**, 521 (1972).
33. W. S. D. Wilcock, *J. Geophys. Res.* **103**, 2585 (1998).
34. E. M. Van Ark et al., *J. Geophys. Res.* **112**, 10.1029/2005JB004210 (2007).
35. D. Glickson, D. Kelley, J. Delaney, *Geochem. Geophys. Geosyst.* **8**, Q06010 (2007).
36. J. Hemley, G. Cygan, J. Fein, G. Robinson, W. Angelo, *Econ. Geol.* **87**, 1 (1992).
37. J. Franklin, H. Gibson, I. Jonasson, A. Galley, in *Economic Geology, 100th Anniversary Volume*, J. W. Hedenquist, J. F. H. Thompson, R. J. Goldfarb, J. P. Richards, Eds. (Society of Economic Geologists, Littleton, CO, 2005), pp. 523–560.
38. M. Hannington, I. Jonasson, P. Herzig, S. Petersen, *Geophys. Monogr.* **91**, 115 (1995).
39. This work was supported by the Swiss National Science Foundation (grant 200020-107955). The authors thank S. Geiger for providing the mesh depicted in Fig. 1 and S. Ingebritsen, P. Weiss, L. Cathles, and three anonymous reviewers for useful discussions as well as careful proofreading of earlier versions of the manuscript.

Neodymium-142 Evidence for Hadean Mafic Crust

Jonathan O'Neil,[1]* Richard W. Carlson,[2] Don Francis,[1] Ross K. Stevenson[3]

Neodymium-142 data for rocks from the Nuvvuagittuq greenstone belt in northern Quebec, Canada, show that some rock types have lower $^{142}Nd/^{144}Nd$ ratios than the terrestrial standard ($\varepsilon^{142}Nd = -0.07$ to -0.15). Within a mafic amphibolite unit, $^{142}Nd/^{144}Nd$ ratios correlate positively with Sm/Nd ratios and produce a ^{146}Sm-^{142}Nd isochron with an age of 4280^{+53}_{-81} million years. These rocks thus sample incompatible-element-enriched material formed shortly after Earth formation and may represent the oldest preserved crustal section on Earth.

The past decade has seen dramatic discoveries concerning the oldest rocks on Earth, with precise zircon ages pushing the terrestrial rock record back beyond 4 billion years ago (Ga) (*1*) and the detrital zircon record to beyond 4.3 Ga (*2*). Because zircon is a rare to non-existent phase in most mafic rocks, prospecting for ancient crust through zircon analysis has focused the search on the more evolved rock types that likely, as today, do not represent the major volume of Earth's crust.

The short-lived ^{146}Sm-^{142}Nd isotopic system [half life ($T_{1/2}$) = 103 million years (My)] has proven useful for investigating the early differentiation of the silicate portion of Earth. Recent measurements of the ^{146}Sm-^{142}Nd system in Eoarchean (4.0 to 3.6 Ga) rocks, primarily from Greenland, show excesses in $^{142}Nd/^{144}Nd$ ratios of 10 to 20 parts per million (ppm) compared to modern terrestrial standards testifying to Earth differentiation events within a few tens of million years of Earth formation (*3*–*7*). The high $^{142}Nd/^{144}Nd$ measured

for these rocks indicate that the Eoarchean crustal rocks were sourced in a mantle with high Sm/Nd ratio. We describe evidence from the Nuvvuagittuq greenstone belt that a complimentary, low Sm/Nd ratio, reservoir is also found in the terrestrial rock record and that these rocks may be the oldest yet discovered on Earth.

The recent discovery of the Nuvvuagittuq greenstone belt in Ungava, Québec, provides a new suite of Eoarchean rocks with which to further our understanding of the early crust-mantle system. The Nuvvuagittuq belt exposes volcanic and metasedimentary rocks in an isoclinal synform refolded into a more open south-plunging synform (Fig. 1) (*8*) and is surrounded by a 3.66-billion-

[1]Earth and Planetary Sciences Department, McGill University, 3450 University Street, Montreal, Quebec, H3A 2A7, Canada. [2]Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington, DC 20015, USA. [3]GEOTOP (Centre de recherche en géochimie et géodynamique), Université du Québec à Montréal, Post Office Box 8888, Succursale Centre-ville, 210, Président-Kennedy Avenue, Montreal, Quebec H3C 3P8, Canada.

*To whom correspondence should be addressed. E-mail: oneil_jo@eps.mcgill.ca

year (Gy) tonalite (*9*, *10*). Geochronological constraints for the belt come mainly from rare felsic bands (0.5 to 1 m in width) composed of plagioclase, biotite, and quartz that have yielded a discordant zircon age of 3817 ± 16 My (*9*). A minimum age of 3750 My has also been obtained from U-Pb ion microprobe analyses of zircons found in a similar lithology (*11*). Although no clear crosscutting relationship has been found, the felsic bands are commonly found within gabbroic sills that may suggest an intrusive nature for these rocks. The dominant lithology of the belt is a cummingtonite-amphibolite referred to as "faux-amphibolite" because of its unusual mineralogical composition in which the dominant amphibole is cummingtonite, in contrast to the hornblende-dominated amphibolites usually found in the Superior Province. The faux-amphibolite is composed of variable amounts of cummingtonite, plagioclase, biotite, and quartz plus or minus garnet (*8*), commonly with compositional layering defined by the alternation of biotite-rich and cummingtonite-rich laminations. These rocks are very heterogeneous and can be almost entirely composed of cummingtonite, giving them a light gray color, whereas some are reddish brown in color and mainly consist of biotite and garnet, with minor amounts of cummingtonite.

The faux-amphibolite is intruded in the western limb of the synform by ultramafic and gabbroic sills. Compared with these relatively undeformed gabbros, two larger gabbroic sills toward the center of the belt display pronounced gneissic textures, suggesting that they may be older than the less-deformed gabbroic sills. The faux-amphibolite in the western limb rarely contains garnet, whereas toward the center of the belt it has abundant garnet and commonly higher Al_2O_3 contents (*8*). The faux-amphibolite was originally interpreted to be a paragneiss (*10*) because of the abundance of garnet and compositional layering. These rocks, however, are more mafic than typical Archean shales and have a basaltic major element composition similar to those of the gabbroic sills but with lower CaO, slightly lower TiO_2, and commonly higher Al_2O_3 contents. These features suggested that they could be highly altered mafic pyroclastites comagmatic with the gabbro sills (*8*). The faux-amphibolite, however, is enriched in light rare earth elements (LREEs), which argues against a direct cogenetic relationship with the gabbros that intrude it, all of which have flat REE patterns (table S1).

Seven samples of the faux-amphibolite yielded low $^{147}Sm/^{144}Nd$ ratios (0.143 to 0.179) with correspondingly low measured $^{143}Nd/^{144}Nd$ ($\varepsilon^{143}Nd$ from -27.4 to -9.7, where $\varepsilon^{143}Nd = [(^{143}Nd/^{144}Nd)_{sample}/(^{143}Nd/^{144}Nd)_{CHUR} - 1] \times 10^4$ and CHUR is chondritic uniform reservoir) and $^{142}Nd/^{144}Nd$ ratios ($\varepsilon^{142}Nd = -0.07$ to -0.15, where $\varepsilon^{142}Nd = [(^{142}Nd/^{144}Nd)_{sample}/(^{142}Nd/^{144}Nd)_{standard}) - 1] \times 10^4$) relative to the terrestrial standard [Fig. 2, table S2, and supporting online material (SOM) text]. Two tonal-

ite samples also show low $^{142}Nd/^{144}Nd$ ($\varepsilon^{142}Nd = -0.12$ to -0.16) as do both felsic band samples ($\varepsilon^{142}Nd = -0.09$ to -0.10), although the felsic bands overlap the terrestrial standard within the external error of 0.06 as determined by the 2σ-population reproducibility of the standards (SOM text). The gabbros have $^{147}Sm/^{144}Nd$ ratios (0.183 to 0.193) and measured $^{143}Nd/^{144}Nd$ ratios ($\varepsilon^{143}Nd = -4.7$ to 0.6) that are just slightly below chondritic values, with $^{142}Nd/^{144}Nd$ ratios overlapping the terrestrial standard ($\varepsilon^{142}Nd = -0.09$ to -0.02).

The Nuvvuagittuq rocks have $^{142}Nd/^{144}Nd$ ratios that fall between the $^{142}Nd/^{144}Nd$ ratios of chondrites and the terrestrial standard [e.g., (*12*)]. This is in contrast to Eoarchean rocks of Greenland that have $^{142}Nd/^{144}Nd$ values higher than the terrestrial standard (*3–7*). Also, unlike the Greenland rocks, the $^{142}Nd/^{144}Nd$ ratios of the gabbros and faux-amphibolite correlate positively with their Sm/Nd ratios, producing a statistically significant slope corresponding to a $^{146}Sm/^{144}Sm$ ratio of 0.00116 ± 0.00049 [mean square weighted deviation (MSWD) = 0.67, error with 95% confidence] and an initial $\varepsilon^{142}Nd = -0.02 \pm 0.15$ relative to the terrestrial standard (Fig. 3). This line is fit assigning a constant ± 6 ppm error for $^{142}Nd/^{144}Nd$ for all samples to represent the external reproducibility of these isotope ratio determinations. For a solar system initial

$^{146}Sm/^{144}Nd = 0.008$ (*13*) at 4567 Ga (*14*), this slope corresponds to an age of 4280^{+53}_{-81} My. Fitting just the faux-amphibolite data provides a slope of 0.0012 ± 0.0011 corresponding to an age of 4286^{+96}_{-370} My with an initial $\varepsilon^{142}Nd = +0.02 \pm 0.32$ relative to the terrestrial standard. Both the tonalites and the felsic bands fall off these correlations to the low Sm/Nd ratio side.

At the whole-rock scale, isochrons need not provide the crystallization ages of the rocks that define the isochron. Because 4.28-Gy-old the isochron shown in Fig. 3 goes through the value of the modern terrestrial mantle, the unfractionated rocks would plot close to the modern mantle value on Fig. 3 regardless of their age. For example, 12 ultramafic to gabbroic samples from one sill cutting the faux-amphibolite give a $^{147}Sm-^{143}Nd$ isochron age of 3840 ± 280 My (MSWD = 3.8) with initial $\varepsilon^{143}Nd = +0.9$ (Fig. 4A), but the data for some of the same samples lie close to the 4.28-Gy isochron in Fig. 3. The old age in Fig. 3 is dictated by the low Sm/Nd ratios of the faux-amphibolite. The faux-amphibolite defines a scattered $^{147}Sm-^{143}Nd$ correlation with an age of 3819 ± 270 My (MSWD = 5.5) and an initial $\varepsilon^{143}Nd = -1.4$. The negative initial $\varepsilon^{143}Nd$ of the faux-amphibolites is unusual for rocks of this age but is consistent with the low $^{142}Nd/^{144}Nd$ ratios measured for these samples.
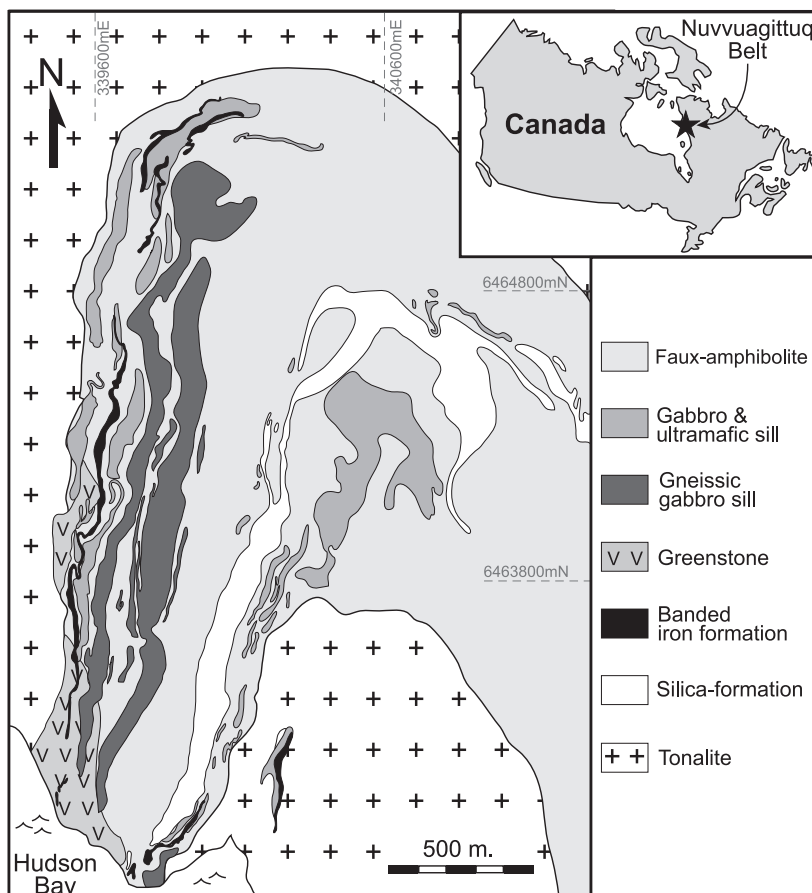
Fig. 1. Simplified geological map of the Nuvvuagittuq greenstone belt. The geology of the belt is described in more detail in (*8*). Coordinates are in universal transverse mercator zone 18 NAD 27.

Sm/Nd ratios as low as those measured in the faux-amphibolite and felsic bands would not produce $^{142}Nd/^{144}Nd$ ratios outside of measurement uncertainty from the terrestrial standard if this parent/daughter fractionation occurred later than ~4.1 to 4.2 Ga. Thus, even if the faux-amphibolite has crystallization ages of 3.8 Gy, the steepness of the $^{142}Nd/^{144}Nd$–Sm/Nd covariation requires that they sample a LREE-enriched material that is at least 4.28 Gy old. This LREE-enriched component could be either older crust that contaminated parental melts like the gabbros or a LREE-enriched mantle source that melted to produce the faux-amphibolite. Regardless of the nature of the LREE-enriched component, its low Sm/Nd ratio must have formed while ^{146}Sm was still extant, and the 4.28-Gy-old isochron provides the best indication of the age of this end member.

An alternate interpretation is that the 4.28-Gy-old isochron indeed dates the formation age of the faux-amphibolite, but this possibility is not supported by the 3.8-Gy-old ^{147}Sm-^{143}Nd age of this unit. For an isochron that passes through the terrestrial mantle point, however, a reduction in Sm/Nd ratio of the faux-amphibolite with the lowest Sm/Nd ratios by only 4.4% caused by metamorphism at 3.8 Ga would rotate a ^{147}Sm-^{143}Nd isochron of 4.28 Gy to 3.8 Gy. Increasing the Sm/Nd ratio of the low–Sm/Nd ratio faux-amphibolite by 4.4% would increase the ^{146}Sm-^{142}Nd isochron age by only 25 My, well within the uncertainty of the data, illustrating the potential of the ^{146}Sm-^{142}Nd system to see through later metamorphic events.

Obviously, other corroborative data would help resolve whether the 4.28-Gy age dates the rocks themselves or an older component involved in their genesis. In spite of attempts to do so, zircons have not yet been found in the faux-amphibolite. Whole-rock Pb isotope data for the faux-amphibolite (table S1) do not define a valid isochron. The best fit line through the $^{206}Pb/^{204}Pb$–$^{207}Pb/^{204}Pb$ data corresponds to an age of 2.4 ± 0.4 Gy, indicative of a late disturbance of the U-Pb system at the whole-rock scale. The faux-amphibolite is crosscut by the gabbro sills and therefore must be older than the gabbros. Although the undeformed gabbros give a ^{147}Sm-^{143}Nd isochron age of 3.84 Gy (Fig. 4A), an isochron constructed from nine samples of the more gneissic, presumably older, gabbros gives an age of 4023 ± 110 My (MSWD = 0.78) with initial $\varepsilon^{143}Nd$ = +1.7 (Fig. 4B). All samples of the faux-amphibolite, except PC-129, yield negative $\varepsilon^{143}Nd$(3.8 Gy) values ranging from –3.2 to –1.0, compared with the mostly positive $\varepsilon^{143}Nd$(3.8 Ga) (–0.2 to +3.1) for the gabbro and ultramafic sills (table S2). When the $\varepsilon^{143}Nd$ values for the faux-amphibolite, except PC-129, are calculated for an age of 4.28 Gy, they range from – 0.3 to +2.3 with an average value of 0.6. This average initial $\varepsilon^{143}Nd$ value is consistent with the mantle value at 4.28 Ga predicted by various depleted mantle evolution models (5, 15, 16), but

Fig. 2. The $^{142}Nd/^{144}Nd$ ratios for the Nuvvuagittuq rocks normalized to the La Jolla standard. Gray solid bar corresponds to the external (6 ppm) error obtained on the terrestrial standard. Error bars for individual samples correspond to either the 2σ-mean internal precision of the mass spectrometer analysis or the 2σ-mean of repeat analysis of the sample on the same mass spectrometer filament load. Solid circle indicates ultramafic sill; open circles, gabbro; diamonds, faux-amphibolite (alternating open and solid diamonds show data for replicate analyses of single samples); squares, tonalite; and triangles (alternating open and solid triangles show data for replicate analyses of single samples), felsic band.
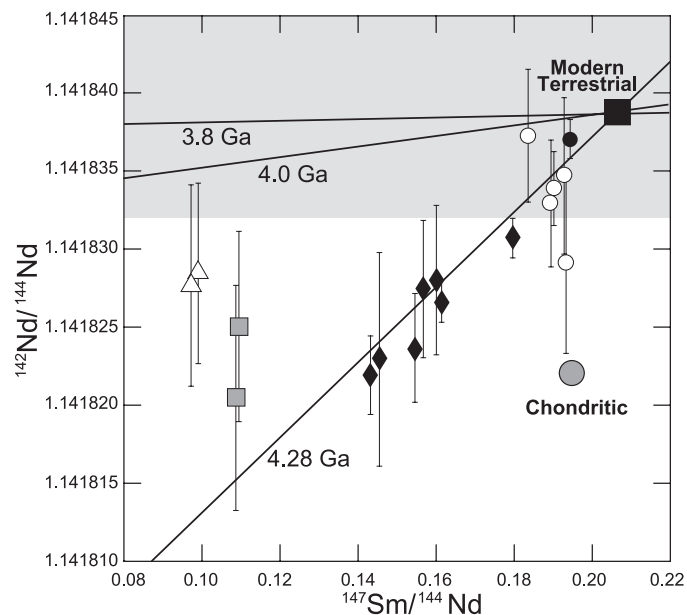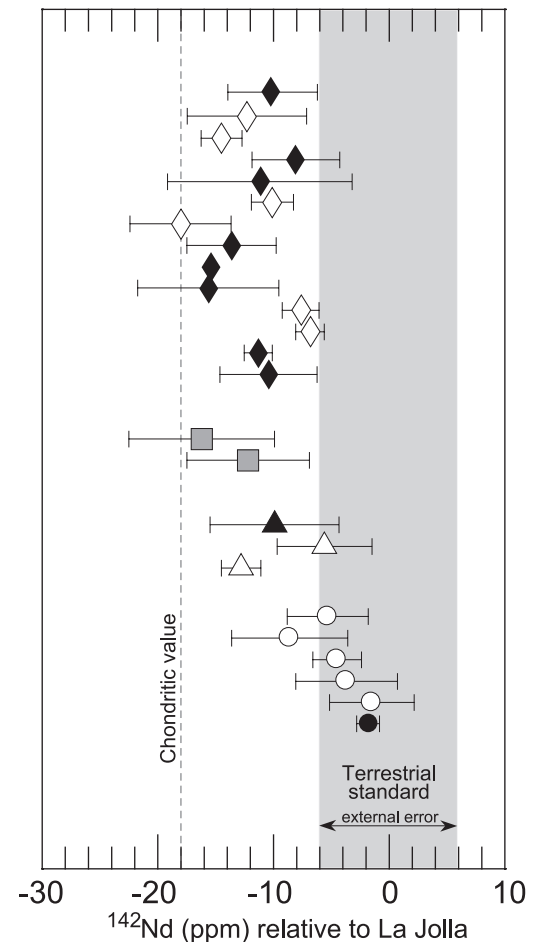
Fig. 3. $^{142}Nd/^{144}Nd$ versus $^{147}Sm/^{144}Nd$ isochron diagram. Symbols as described for Fig. 2. Only the average value of replicate analyses is plotted on this figure. The horizontal gray band shows the ± 6 ppm external precision obtained on the terrestrial standard. Error bars on individual samples are either the 2σ mean of multiple analyses or the 2σ mean of the individual mass spectrometer run for samples run only once. The best fit line through the faux-amphibolite and gabbro data corresponding to an age of 4.28 Gy is shown, as are 3.8- and 4.0-Gy isochrons for reference. The gray circle shows the average value measured for ordinary and enstatite chondrites (12).

Fig. 4. ^{147}Sm-^{143}Nd isochron diagrams. (**A**) Ultramafic to gabbroic samples from one differentiated sill. (**B**) Samples from the gabbro sills that show a strong gneissic metamorphic texture.

A

Age = 3840 ± 280 Ma
Initial ^{143}Nd/^{144}Nd =0.50769 ± 0.00036
MSWD = 3.8

B

Age = 4023 ± 110 Ma
Initial ^{143}Nd/^{144}Nd =0.50749 ± 0.00014
MSWD = 0.78
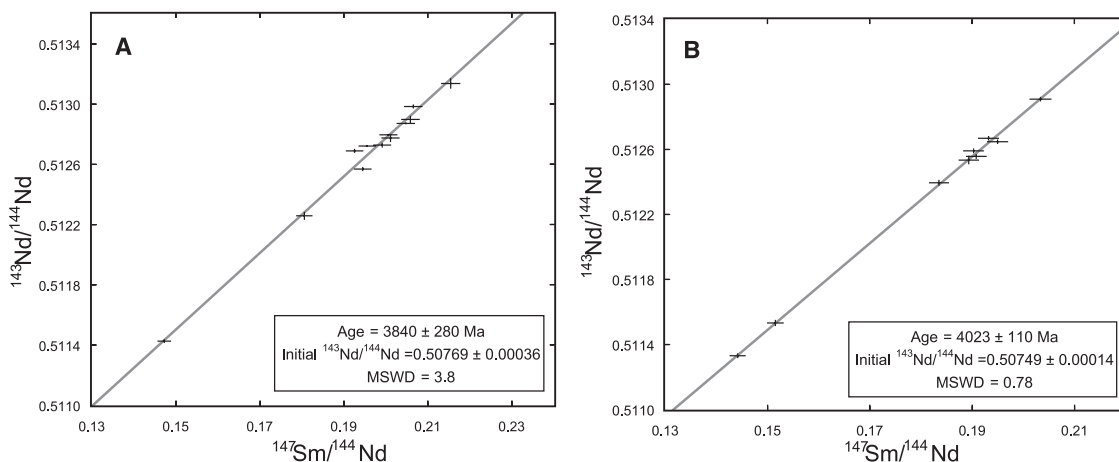
perhaps not as depleted as suggested by the Greenland data (7, 17, 19). As a result, the calculated ^{143}Nd depleted-mantle model ages (T_{DM}) for the faux-amphibolite, except PC-129, range from 4.1 to 4.4 Ga, consistent with the age suggested by ^{142}Nd systematics and in contrast to the 3.2 to 3.6 Ga T_{DM} values of the gabbros and sample PC-129 (table S2).

Whether or not the faux-amphibolite is 4.28 Gy old, its compositional characteristics may provide clues to the process of crust formation in the Hadean (>4.0 Ga). The basaltic major and compatible (e.g., Ni) trace element composition of the faux-amphibolite is consistent with derivation from a peridotitic mantle. Compared to the gabbros and to modern mid-ocean ridge basalts, the most unusual compositional characteristic of the faux-amphibolite is its low Ca content, high K and Rb contents, and LREE enrichment. Because elements like K and Rb are easily affected by alteration, however, it is unclear whether these are magmatic features of the faux-amphibolite. The LREE enrichment could reflect relatively low degrees of mantle melting, but this explanation is not supported by the relatively low concentration of elements such as Ti and Nb in the faux-amphibolite. The high LREE to Nb ratios of the faux amphibolite, however, is similar to that of modern calc-alkaline melts produced in convergent margin settings. The Hadean crust, represented by the faux-amphibolite, was intruded at 4.0 and 3.8 Ga by gabbro and ultramafic sills that have the ^{143}Nd and ^{142}Nd isotopic composition of the depleted mantle at the time of their intrusion. The low ^{142}Nd/^{144}Nd ratios of tonalites and felsic bands that were emplaced between 3.8 and 3.6 Ga, well after ^{146}Sm was extinct (9–11), suggest that they formed by the partial melting of the faux-amphibolite.

References and Notes
1. S. A. Bowring, I. S. Williams, *Contrib. Mineral. Petrol.* **134**, 3 (1999).
2. S. A. Wilde, J. W. Valley, W. H. Peck, C. M. Graham, *Nature* **409**, 175 (2001).
3. M. Boyet *et al.*, *Earth Planet. Sci. Lett.* **214**, 427 (2003).
4. G. Caro, B. Bourdon, J. Birck, S. Moorbath, *Nature* **423**, 428 (2004).
5. M. Boyet, R. W. Carlson, *Earth Planet. Sci. Lett.* **250**, 254 (2006).
6. G. Caro, B. Bourdon, J. Birck, S. Moorbath, *Geochim. Cosmochim. Acta* **70**, 164 (2006).
7. V. C. Bennett, A. D. Brandon, A. P. Nutman, *Science* **318**, 1907 (2007).
8. J. O'Neil *et al.*, in *In Earth's Oldest Rocks*, M. van Kranendonk, R. H. Smithies, V. C. Bennett, Eds. (Elsevier, Amsterdam, 2007), pp. 219–250.
9. J. David, L. Godin, R. K. Stevenson, J. O'Neil, D. Francis, *Geol. Soc. Am. Bull.*, in press (available at www.gsajournals.org/perlserv/?request=get-abstract&doi=10.1130%2FB26369.1).
10. M. Simard, M. Parent, J. David, K. N. M. Sharma, "Géologie de la région de la rivière Innuksuac (34K et 34L)" (Ministère des Ressources naturelles, RG 2002-10, Québec, Canada, 2003).
11. N. L. Cates, S. J. Mojzsis, *Earth Planet. Sci. Lett.* **255**, 9 (2007).
12. M. Boyet, R. W. Carlson, *Science* **309**, 576 (2005); published online 16 June 2005 (10.1126/science.1113634).
13. G. W. Lugmair, S. J. G. Galer, *Geochim. Cosmochim. Acta* **56**, 1673 (1992).
14. Y. Amelin, A. N. Krot, I. D. Hutcheon, A. A. Ulyanov, *Science* **297**, 1678 (2002).
15. D. J. DePaolo, *Nature* **291**, 193 (1981).
16. S. L. Goldstein, R. K. O'Nions, P. J. Hamilton, *Earth Planet. Sci. Lett.* **70**, 221 (1984).
17. V. C. Bennett, A. P. Nutman, M. T. McCulloch, *Earth Planet. Sci. Lett.* **119**, 299 (1993).
18. M. T. McCulloch, V. C. Bennett, *Geochim. Cosmochim. Acta* **58**, 4717 (1994).
19. M. T. McCulloch, V. C. Bennett, in *The Earth's Mantle*, I. Jackson, Ed. (Cambridge Univ. Press, Cambridge, 1998), pp. 127–158.
20. This research was supported by National Science and Engineering Research Council of Canada (NSERC) Discovery grants to D.F. (RGPIN 7977-00). We thank the municipality of Inukjuak and the Pituvik Landholding Corporation for permission to work on their territory; J. Mina, M. Carroll, V. Inukpuk Morkill, R. Kasudluak, and J. Williams for their hospitality and support; and M. Horan and T. Mock for analytical support. The Department of Terrestrial Magnetism Thermo-Fisher Scientific Triton (Bremen, Germany) was purchased with partial support from the NSF grant EAR-0320589.

Infants' Perseverative Search Errors Are Induced by Pragmatic Misinterpretation

József Topál,[1]* György Gergely,[1,2] Ádám Miklósi,[3] Ágnes Erdőhegyi,[3] Gergely Csibra[2,4]

Having repeatedly retrieved an object from a location, human infants tend to search the same place even when they observe the object being hidden at another location. This perseverative error is usually explained by infants' inability to inhibit a previously rewarded search response or to recall the new location. We show that the tendency to commit this error is substantially reduced (from 81 to 41%) when the object is hidden in front of 10-month-old infants without the experimenter using the communicative cues that normally accompany object hiding in this task. We suggest that this improvement is due to an interpretive bias that normally helps infants learn from demonstrations but misleads them in the context of a hiding game. Our finding provides an alternative theoretical perspective on the nature of infants' perseverative search errors.

Human infants' abilities for understanding the physical world are often tested in hide-and-search tasks. First demonstrated by Piaget (1), the perseverative search error (some-

times called the A-not-B error) is a well-known and robust mistake that infants close to 1 year of age normally commit. In the standard A-not-B task, a demonstrator repeatedly places an object

under one (A) of two opaque containers (A and B) in full view of the infant. After each hiding event, the infant is allowed to retrieve the object. This is followed by test trials where the demonstrator places the object under container B and allows the infant to search for it. Despite just having seen the object being hidden at the new B location, infants between 8 and 12 months of age frequently look for it under container A where it had been previously hidden. This perseverative search error continues to be of theoretical interest for researchers of cognitive development (2–6).

A wide range of explanations have been proposed to account for this response bias. According to Piaget's original hypothesis (1), the A-not-B error reflects young infants' as yet incomplete comprehension of object permanence. Piaget believed that the infant conceives the appearance of the object under container A to be an inherent consequence of the search response itself. More recent accounts of the perseverative bias have focused on the motor response involved in searching at location A, which has been primed during its repeated execution after the initial hiding trials. In these accounts, the A-not-B error is usually ascribed to a deficit in inhibitory control over a previously rewarded motor response (7) or to constraints on short-term memory (8), or both (9). Alternatively, the perseverative response has been seen as driven by a response bias established in the visuomotor response execution system during repeated A-trials (10). Others point out that simply observing another person reaching to location A repeatedly is in itself sufficient to elicit the A-not-B error. In this view, infants' errors do not reflect their difficulty with response inhibition,

but are due to an attentional bias to the location where the previously observed manual responses have been directed (11). A more recent explanation suggests that observing repeated hiding events at location A leads to automatic motor simulation (covert imitation) of the action through the activation of the mirror neuron system (12).

In contrast to the focus of such accounts on infants' repeated responses directed at container A, we have examined the perseverative error from a different perspective by exploring the potential role of the communicative demonstration context of the task. The A-not-B task normally involves face-to-face interaction, in which object hiding is accompanied by the demonstrator's ostensive and referential signals [such as eye contact, infant-directed speech, addressing the baby by name, and pointing at and/or looking back and forth between the hiding location and the infant (13)]. Recent findings indicate that ostensive-referential communicative signals can play an interpretation-modulating role, leading to selective encoding of different aspects of action demonstrations in social learning tasks [e.g., (14–17)]. Csibra and Gergely (13, 18) hypothesized that ostensive signals induce a receptive "pedagogical learning stance" in the infant, involving a built-in interpretive bias of generalizability. This bias assumes that ostensively communicated manifestations are more likely to convey semantic or generic information about the referent than episodic information that obtains only in the here-and-now.

The hiding events in the standard A-not-B task can be interpreted both as indicating episodic information about the referent's current location ("the target object is now under container A") and as communicating information about some generalizable property of the referent kind (e.g., "this type of object is usually found in container A"). We hypothesized that in the A-not-B paradigm, the interpretive bias of generalizability may result in a pragmatic misinterpretation of the object-hiding actions as potential teaching demonstrations. As a result, the infant would tend to

infer and learn some generalizable information, such as "this kind of object is to be found in container A" or "we keep toys in container A." According to this hypothesis, misinterpretation of the ostensively communicated hiding events leads infants to commit the perseverative search error during B test trials. We therefore predicted that in a noncommunicative action observation condition, which lacks ostensive signals but provides experience with repeated motor search responses directed at container A, the perseverative search error should be reduced.

To test this hypothesis, we examined infants' object search behavior in the A-not-B task while varying the presence or absence of the social-communicative context of the hiding events. Three groups of 10-month-old infants (14 in each) were tested. In the ostensive-communicative context (OC) (Fig. 1A), the demonstrator established eye contact with the baby, smiling at and addressing him or her in infant-directed speech (saying "Hello baby, look here!"). Then she repeatedly hid a toy object under container A while shifting her eye gaze back and forth between the infant and the container to direct and share the infant's attention toward the object-hiding action. In the noncommunicative context (NC) (Fig. 1B), the demonstrator's face and torso were oriented 90° away from the infant and, while her hands were just as visible during the repeated hiding actions as in the OC condition, she never looked at or communicated with the infant in any way while hiding the object. In the nonsocial context (NS) (Fig. 1C), the demonstrator acted from behind a curtain and only the object's movements were visible to the infant. In each condition, after a 4-s delay following the hiding events, the demonstrator slid the cardboard sheet with the two containers closer to the infant and then waited until a search response was executed. The toy was hidden four times at the first location (A-trials), then three times at the other location (B-trials) (19).

We analyzed the proportion of correct responses in both the A- and B-trials, as well as the

[1]Research Institute for Psychology, Hungarian Academy of Sciences, Budapest H-1132, Hungary. [2]Department of Philosophy, Central European University, Budapest H-1051, Hungary. [3]Department of Ethology, Eötvös University, Budapest H-1117, Hungary. [4]School of Psychology, Birkbeck, University of London, London WC1 E7HX, UK.

*To whom correspondence should be addressed. E-mail: topaljozsef@gmail.com

Fig. 1. Experimental arrangement in the three hiding contexts: (**A**) ostensive-communicative task, (**B**) noncommunicative task, and (**C**) nonsocial task.

Ostensive-communicative   Noncommunicative   Nonsocial
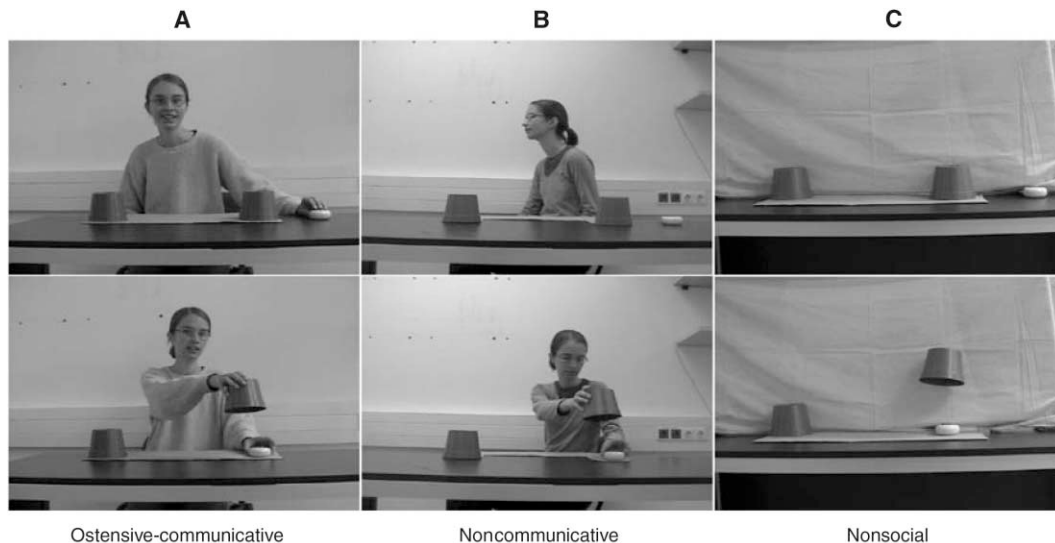
Fig. 2. Proportion of correct searches (mean + SE) in A- and B-trials as a function of the hiding context. The 10-month-old infants received four A-trials followed by three B-trials. ***$P < 0.0001$, *$P < 0.05$.
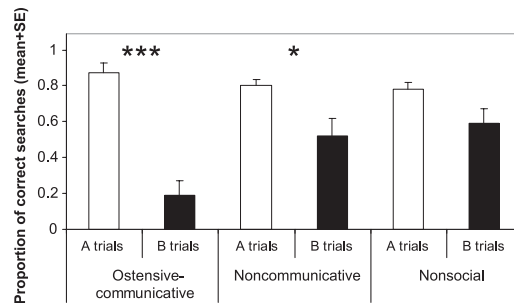
Table 1. Number of infants in the three different hiding contexts (14 in each group) committing at most one or more search errors (searching at location A) in the three B-trials.

Context	0 or 1 error	2 or 3 errors
Ostensive-communicative	2	12
Noncommunicative	8	6
Nonsocial	9	5

number of infants who committed the A-not-B error (searched more than once at location A in the B-trials) as a function of the hiding context. We found that the magnitude of the A-not-B error was considerably smaller in the noncommunicative and nonsocial conditions than in the traditional ostensive-communicative context (Fig. 2 and fig. S1). A two-way analysis of variance on the proportion of correct responses, with phase (A- versus B-trials) and hiding context (OC, NC, NS) as factors, showed more correct responses in the A-trials than in the B-trials ($F_{1,39} = 49.376$, $P < 0.0001$) and a significant interaction between these factors ($F_{2,39} = 8.041$, $P = 0.001$). This interaction was because the change of the proportion of correct searches from the A- to the B-trials differed across contexts. Whereas in the OC context the initial success rate of 0.88 in the A-trials dropped to 0.19 in the B-trials ($t_{13} = 8.917$, $P < 0.0001$), the drop was much smaller in the NC context (0.80 to 0.52, $t_{13} = 2.536$, $P = 0.025$) and was not statistically significant in the NS context (0.78 to 0.59, $t_{13} = 1.96$, $P = 0.072$). In addition, although the infants were similarly successful during the A-trials in all contexts ($F_{2,39} = 1.525$, $P = 0.23$), their search performance differed significantly across contexts in the B-trials ($F_{2,39} = 6.660$, $P = 0.005$). In this latter case, post hoc pairwise comparisons (Tukey-Cramer test) showed that infants searched the least correctly in the OC context (OC versus NO, $P < 0.05$; OC versus NS, $P < 0.01$).

Comparison of the number of infants committing the A-not-B error (Table 1) indicated a significant difference between demonstration conditions ($\chi^2 = 8.265$, $P = 0.016$). After having witnessed ostensive-communicative hiding demonstrations during the A-trials, 86% of the infants displayed the perseverative error during the B-trials. In contrast, the majority of infants in the other two contexts (NC, 57%;

NS, 64%) did not show a perseverative response pattern.

These results are not compatible with the currently widely accepted explanations for the A-not-B perseverative response bias, which attribute this robust developmental phenomenon to the dominance (and lack of inhibition) of the prepotent motor search response. Our results also challenge recent proposals that the motor priming of the prepotent response can be induced by simply observing the manual hiding actions directed at location A, mediated by the mirror neuron system (*12*), because the NC and OC contexts provided the same amount of visual (as well as motor) experience of the repeated manual hiding actions directed at container A.

Rather, we suggest that our results can be explained by the theory of natural pedagogy (*13, 14*), which proposes a special interpretation-modulating role for ostensive-referential signals in early social learning. The action demonstrations of the A-not-B paradigm can be interpreted either as a hide-and-search game, presenting the infant with episodic (here-and-now) information about the whereabouts of the object (correct interpretation), or as a kind of teaching session that conveys generalizable information about properties of the objects (toys or containers) for the infant to learn (incorrect interpretation). We propose that it is this latter kind of interpretation—mistakenly established during the ostensively demonstrated A-trials—that remains dominant during the B-trials, leading to the erroneous perseverative search responses.

This conclusion does not invalidate the contribution of other cognitive factors to the A-not-B error identified by earlier studies. In our study, the perseverative error was reduced but did not completely disappear in the NC and NS contexts, which suggests that infants' search behavior also depends on their inhibitory, information processing, and memory skills (*20–21*). Had we not inserted a 4-s delay between hiding and searching, infants would have been likely to search for the toy at the correct location (*22, 23*). Thus, the decay of the accessibility of the short-term memory of target location for search actions is a necessary component of the search error. Similarly, the ostensive hiding demonstration may have generated higher cognitive load to be overcome in search behavior than did the NC and NS contexts (*22*). Although these accounts could explain why infants did not search at the correct location, they

predict random search rather than perseveration. However, it was only in the NS and NC conditions that infants' search pattern was close to 50%; in the OC condition, they tended to search more often at location A. The theory of natural pedagogy offers an explanation for this baffling tendency to perseverate: The communicative demonstration during the A-trials generates a semantic (and potentially long-term) memory trace that biases infants to search at the old location when they no longer have access to the decayed memory trace of the current location or when their information-processing capacity is overloaded. In addition, the primacy effect of semantic learning (as opposed to the recency effect on episodic memory) made it difficult for the infants to relearn the new location in B-trials.

Human infants are highly social creatures (*24*) who cannot help but interpret the ostensive communicative signals directed to them. Although such a disposition prepares them to efficiently learn from adults, in certain situations (e.g., the A-not-B task) it can also misguide their performance. Our demonstration of the social communicative determinants of infants' early tendency for perseveration in motor search tasks provides independent support for our general proposal (*13*) that sensitivity to ostensive-referential communication is a basic evolutionary adaptation that is fundamental to the emergence of human social cognition.

References and Notes
1. J. Piaget, *The Construction of Reality in the Child* (Basic Books, New York, 1954).
2. P. L. Harris, in *Infant Development*, A. Slater, G. Bremner, Eds. (Erlbaum, Hove, UK, 1989), pp. 103–122.
3. H. M. Wellman, D. Cross, K. Bartsch, *Monogr. Soc. Res. Child Dev.* **51**, 3 (1986).
4. S. Marcovitch, P. D. Zelazo, *Child Dev.* **70**, 1297 (1999).
5. E. Thelen, G. Schöner, C. Scheier, L. B. Smith, *Behav. Brain Sci.* **24**, 1 (2001).
6. T. Ruffman, L. Slade, J. C. Sandino, A. Fletcher, *Child Dev.* **76**, 122 (2005).
7. A. Diamond, *Child Dev.* **56**, 868 (1985).
8. E. M. Cummings, E. L. Bjork, *Mem. Cognit.* **12**, 1 (1984).
9. A. Diamond, L. Cruttenden, D. Neiderman, *Dev. Psychol.* **30**, 192 (1994).
10. L. B. Smith, E. Thelen, R. Titzer, D. McLin, *Psychol. Rev.* **106**, 235 (1999).
11. T. Ruffman, L. Langman, *Infant Behav. Dev.* **25**, 237 (2002).
12. M. R. Longo, B. I. Berthental, *Infancy* **10**, 43 (2006).
13. G. Csibra, G. Gergely, in *Processes of Change in Brain and Cognitive Development*, Y. Munakata, M. Johnson, Eds. (Oxford Univ. Press, Oxford, 2006), pp. 249–274.
14. G. Gergely, K. Egyed, I. Király, *Dev. Sci.* **10**, 139 (2007).
15. A. Brugger, L. A. Lariviere, D. L. Mumme, E. W. Bushnell, *Child Dev.* **78**, 806 (2007).
16. M. Nielsen, *Dev. Psychol.* **42**, 555 (2006).
17. Y. M. D. Yoon, M. H. Johnson, G. Csibra, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13690 (2008).
18. G. Gergely, G. Csibra, in *Roots of Human Sociality: Culture, Cognition, and Human Interaction*, N. J. Enfield, S. C. Levinson, Eds. (Berg, Oxford, 2006), pp. 229–255.
19. See supporting material on *Science* Online.
20. A. Diamond, *Ann. N.Y. Acad. Sci.* **608**, 267 (1990).
21. R. Keen, R. L. Caricco, M. R. Sylvia, N. E. Berthier, *Dev. Sci.* **6**, 221 (2003).
22. A. Diamond, P. S. Goldman-Rakic, *Exp. Brain Res.* **74**, 24 (1989).
23. A. Ahmed, T. Ruffman, *Dev. Psychol.* **34**, 441 (1998).

24. E. Herrmann, J. Call, M. V. Hernàndez-Lloreda, B. Hare, M. Tomasello, *Science* **317**, 1360 (2007).
25. We thank J. Bognár for her assistance in data collection, and V. Southgate and J. Watson for helpful comments. Supported by the Hungarian Scientific Research Fund (T049615), the EU6 Framework Programme (NEUROCOM grant 12738, EDICI grant 12929), and the Bolyai Foundation of the Hungarian Academy of Sciences.

Antigen Recognition by Variable Lymphocyte Receptors

Byung Woo Han,[1,2] Brantley R. Herrin,[3] Max D. Cooper,[3] Ian A. Wilson[1,2]*

Variable lymphocyte receptors (VLRs) rather than antibodies play the primary role in recognition of antigens in the adaptive immune system of jawless vertebrates. Combinatorial assembly of leucine-rich repeat (LRR) gene segments achieves the required repertoire for antigen recognition. We have determined a crystal structure for a VLR-antigen complex, VLR RBC36 in complex with the H-antigen trisaccharide from human blood type O erythrocytes, at 1.67 angstrom resolution. RBC36 binds the H-trisaccharide on the concave surface of the LRR modules of the solenoid structure where three key hydrophilic residues, multiple van der Waals interactions, and the highly variable insert of the carboxyl-terminal LRR module determine antigen recognition and specificity. The concave surface assembled from the most highly variable regions of the LRRs, along with diversity in the sequence and length of the highly variable insert, can account for the recognition of diverse antigens by VLRs.

In the lamprey and hagfish, the only surviving jawless vertebrates, variable lymphocyte receptors (VLRs) play the major role in recognition of foreign antigens (*1*, *2*). In contrast to the variable, diverse, and joining gene segments (VDJs) of immunoglobulins in jawed vertebrates, the jawless vertebrates have solved the receptor diversity problem by somatic DNA rearrangement of diverse leucine-rich repeat (LRR) modules into incomplete *vlr* genes. The resulting mature *vlr* genes encode an N-terminal LRR capping region (LRRNT), the first LRR (LRR1), up to seven 24-residue variable LRRs (LRRVs) (*3*), a terminal or end LRRV (LRRVe), a connecting peptide (CP), a C-terminal LRR capping region (LRRCT), and a threonine/proline-rich stalk region that connects the protein to a glycosylphosphatidylinositol (GPI) anchor and a hydrophobic tail (Fig. 1A) (*1*, *2*, *4*).

From these somatic gene rearrangements, a potential repertoire of about 10^{14} unique VLRs has been estimated (*2*), which compares favorably with the equivalent diversity attainable through VDJ recombination in antibodies. Different numbers and combinations of LRR modules, coupled with amino acid sequence variation in the LRR segments, thereby contribute to VLR diversity. The LRR repeats form a curved solenoid, as in Toll-like receptors (TLRs) (*5*, *6*), and its concave surface has been suggested as the antigen-binding site from evolutionary, sequence, and mutational analyses (*2*, *7*, *8*). Crystal structures of three unliganded hagfish VLRs with different numbers of LRRV modules have been determined (*7*), whereas antigen-binding specificity [erythrocyte H-trisaccharide (*9*) and *Bacillus* collagen-like protein of *B. anthracis* (BclA) (*8*)] has been reported only for lamprey VLRs. However, the mode of antigen recognition has not yet been determined in either system, nor has it been shown whether complementarity-determining region (CDR) equivalents are present in VLRs that would endow them with specificity and affinity for any given antigen, as for antibodies.

We determined the crystal structure of the VLR RBC36 ectodomain (ECD) in complex with the H-trisaccharide derived from the H-antigen of human blood group O erythrocytes at 1.67 Å resolution by molecular replacement, using our lamprey VLR2913 crystal structure [Protein Data Bank (PDB) ID 2R9U]. Lampreys were previously shown to produce high-titer agglutinins against the H-antigens of human O erythrocytes (*10*, *11*). When lampreys were immunized with human blood group O erythrocytes, they elicited VLRs that recognize the dominant H-trisaccharide antigen on Chinese hamster ovary cells transfected with 1,2-fucosyltransferase (*9*). H-antigens contain the characteristic disaccharide α-L-Fuc*p*-(1→2)-β-D-Gal*p*-OR, where R is glycoprotein or glycolipid (*12*). The type II H-antigen trisaccharide, α-L-Fuc*p*-(1→2)-β-D-Gal*p*-(1→4)-β-D-GlcNac*p*-OH, was

Fig. 1. Overall architecture of the VLR RBC36-ECD in complex with the H-trisaccharide. (**A**) Schematic diagram of RBC36. Regions from left to right: signal peptide (SP), N-terminal LRR (LRRNT), five variable LRRs (LRR1, LRRVs), connecting peptide (CP), C-terminal LRR (LRRCT), threonine/proline-rich stalk region, GPI anchor, and hydrophobic tail. (**B**) Ribbon diagram of RBC36-ECD in complex with H-trisaccharide. LRRNT, LRRs, and LRRCT are colored blue, green, and red, respectively. Carbons, nitrogens, and oxygens of the H-trisaccharide are colored yellow, blue, and red, respectively. Disulfide bridges are shown in orange. Green dotted lines represent hydrogen bonds; black dotted lines indicate hydrophobic effects. (**C**) View rotated 90° from (B) that highlights the continuous β sheet and the H-trisaccharide binding site on the concave surface.

[1]Department of Molecular Biology, Scripps Research Institute, La Jolla, CA 92037, USA. [2]Skaggs Institute for Chemical Biology, Scripps Research Institute, La Jolla, CA 92037, USA. [3]Emory Vaccine Center and Department of Pathology and Laboratory Medicine, Emory University, 1462 Clifton Road NE, Atlanta, GA 30322, USA.

*To whom correspondence should be addressed. E-mail: wilson@scripps.edu

Antigen Recognition by Variable Lymphocyte Receptors

Byung Woo Han, *et al.*
Science **321**, 1834 (2008);
DOI: 10.1126/science.1162484

24. E. Herrmann, J. Call, M. V. Hernàndez-Lloreda, B. Hare, M. Tomasello, *Science* **317**, 1360 (2007).
25. We thank J. Bognár for her assistance in data collection, and V. Southgate and J. Watson for helpful comments. Supported by the Hungarian Scientific Research Fund (T049615), the EU6 Framework Programme (NEUROCOM

Antigen Recognition by Variable Lymphocyte Receptors

Byung Woo Han,[1,2] Brantley R. Herrin,[3] Max D. Cooper,[3] Ian A. Wilson[1,2]*

Variable lymphocyte receptors (VLRs) rather than antibodies play the primary role in recognition of antigens in the adaptive immune system of jawless vertebrates. Combinatorial assembly of leucine-rich repeat (LRR) gene segments achieves the required repertoire for antigen recognition. We have determined a crystal structure for a VLR-antigen complex, VLR RBC36 in complex with the H-antigen trisaccharide from human blood type O erythrocytes, at 1.67 angstrom resolution. RBC36 binds the H-trisaccharide on the concave surface of the LRR modules of the solenoid structure where three key hydrophilic residues, multiple van der Waals interactions, and the highly variable insert of the carboxyl-terminal LRR module determine antigen recognition and specificity. The concave surface assembled from the most highly variable regions of the LRRs, along with diversity in the sequence and length of the highly variable insert, can account for the recognition of diverse antigens by VLRs.

In the lamprey and hagfish, the only surviving jawless vertebrates, variable lymphocyte receptors (VLRs) play the major role in recognition of foreign antigens (*1*, *2*). In contrast to the variable, diverse, and joining gene segments (VDJs) of immunoglobulins in jawed vertebrates, the jawless vertebrates have solved the receptor diversity problem by somatic DNA rearrangement of diverse leucine-rich repeat (LRR) modules into incomplete *vlr* genes. The resulting mature *vlr* genes encode an N-terminal LRR capping region (LRRNT), the first LRR (LRR1), up to seven 24-residue variable LRRs (LRRVs) (*3*), a terminal or end LRRV (LRRVe), a connecting peptide (CP), a C-terminal LRR capping region (LRRCT), and a threonine/proline-rich stalk region that connects the protein to a glycosylphosphatidylinositol (GPI) anchor and a hydrophobic tail (Fig. 1A) (*1*, *2*, *4*).
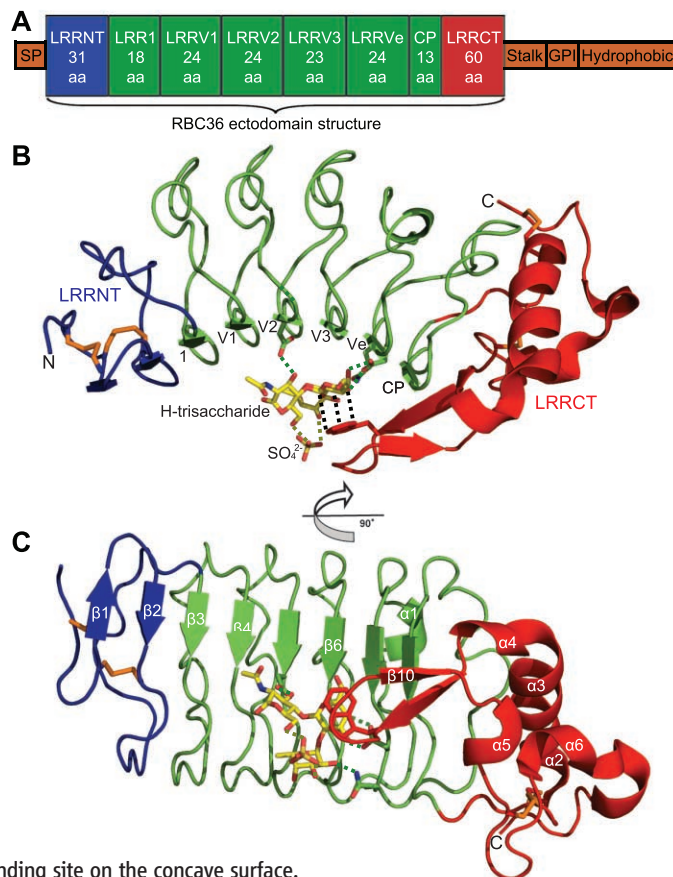
From these somatic gene rearrangements, a potential repertoire of about 10^{14} unique VLRs has been estimated (*2*), which compares favorably with the equivalent diversity attainable through VDJ recombination in antibodies. Different numbers and combinations of LRR modules, coupled with amino acid sequence variation in the LRR segments, thereby contribute to VLR diversity. The LRR repeats form a curved solenoid, as in Toll-like receptors (TLRs) (*5*, *6*), and its concave surface has been suggested as the antigen-binding site from evolutionary, sequence, and mutational analyses (*2*, *7*, *8*). Crystal structures of three unliganded hagfish VLRs with different numbers of LRRV modules have been determined (*7*), whereas antigen-binding specificity [erythrocyte H-trisaccharide (*9*) and *Bacillus* collagen-like protein of *B. anthracis* (BclA) (*8*)] has been reported only for lamprey VLRs. However, the mode of antigen recognition has not yet been determined in either system, nor has it been shown whether complementarity-determining region (CDR) equivalents are present in VLRs that would endow them with specificity and affinity for any given antigen, as for antibodies.

We determined the crystal structure of the VLR RBC36 ectodomain (ECD) in complex with the H-trisaccharide derived from the H-antigen of human blood group O erythrocytes at 1.67 Å resolution by molecular replacement, using our lamprey VLR2913 crystal structure [Protein Data Bank (PDB) ID 2R9U]. Lampreys were previously shown to produce high-titer agglutinins against the H-antigens of human O erythrocytes (*10*, *11*). When lampreys were immunized with human blood group O erythrocytes, they elicited VLRs that recognize the dominant H-trisaccharide antigen on Chinese hamster ovary cells transfected with 1,2-fucosyltransferase (*9*). H-antigens contain the characteristic disaccharide α-L-Fuc*p*-(1→2)-β-D-Gal*p*-OR, where R is glycoprotein or glycolipid (*12*). The type II H-antigen trisaccharide, α-L-Fuc*p*-(1→2)-β-D-Gal*p*-(1→4)-β-D-GlcNac*p*-OH, was

Fig. 1. Overall architecture of the VLR RBC36-ECD in complex with the H-trisaccharide. (**A**) Schematic diagram of RBC36. Regions from left to right: signal peptide (SP), N-terminal LRR (LRRNT), five variable LRRs (LRR1, LRRVs), connecting peptide (CP), C-terminal LRR (LRRCT), threonine/proline-rich stalk region, GPI anchor, and hydrophobic tail. (**B**) Ribbon diagram of RBC36-ECD in complex with H-trisaccharide. LRRNT, LRRs, and LRRCT are colored blue, green, and red, respectively. Carbons, nitrogens, and oxygens of the H-trisaccharide are colored yellow, blue, and red, respectively. Disulfide bridges are shown in orange. Green dotted lines represent hydrogen bonds; black dotted lines indicate hydrophobic effects. (**C**) View rotated 90° from (B) that highlights the continuous β sheet and the H-trisaccharide binding site on the concave surface.

[1]Department of Molecular Biology, Scripps Research Institute, La Jolla, CA 92037, USA. [2]Skaggs Institute for Chemical Biology, Scripps Research Institute, La Jolla, CA 92037, USA. [3]Emory Vaccine Center and Department of Pathology and Laboratory Medicine, Emory University, 1462 Clifton Road NE, Atlanta, GA 30322, USA.

*To whom correspondence should be addressed. E-mail: wilson@scripps.edu

used as the antigen in the crystal structure with the RBC36-ECD (Fig. 1, Fig. 2, A and B, and fig. S1) (13).

Lamprey RBC36-ECD (residues 22 to 238, lacking the N-terminal signal sequence) forms a horseshoe-shaped assembly that is more abbreviated and crescent-shaped relative to TLRs. This assembly consists of an LRRNT, an 18-residue LRR1, three LRRVs, an LRRVe, a CP, and an LRRCT, all of which adopt a right-handed solenoidal structure, except for LRRCT. The inner, concave surface is formed from eight β strands (two from LRRNT, five from LRRs, and one from CP), which assemble into a continuous β sheet. The convex (outer) surface is composed of the more diverse secondary structure elements, including loops of varying length, one α helix, and six 3_{10} helices (Fig. 1). Lamprey RBC36-ECD contains five canonical LRR-signature motifs, $xL^2xxL^5xxL^8xL^{10}xxN^{13}Q^{14}L^{15}xxL^{18}P^{19}xG^{21}V^{22}F^{23}D^{24}$ [where L represents obligate hydrophobic residues—which, for RBC36-ECD, include leucine (most prevalent), isoleucine, or methionine—and N, Q, P, G, F, and D are conserved asparagine, glutamine, proline, glycine, phenylalanine, and aspartic acid residues, respectively] (Fig. 2D) (4, 14). The side chains of nine conserved residues in each LRR (at relative positions 2, 5, 8, 10, 13, 15, 18, 22, and 23) assemble within the solenoidal structure and form a tight hydrophobic core that laterally stabilizes the repeating LRR modules (Fig. 2E). In RBC36-ECD, LRRNT comprises residues 22 to 52, in which Thr^{32}, Val^{33}, and Asp^{34} initiate an antiparallel β strand that extends the continuous parallel β sheet, and LRRNT and LRRCT cover the exposed edges of the hydrophobic core of the solenoidal LRR structure, as observed in other LRR proteins, including TLRs (5, 6). In LRRNT, the characteristic four-cysteine motif (Cx_nCxCx_nC) forms two sets of disulfides (Cys^{22} to Cys^{28} and Cys^{26} to Cys^{35}), whereas in LRRCT, a similar motif ($CxCx_nCx_nC$) gives rise to disulfides Cys^{182} to Cys^{217} and Cys^{184} to Cys^{237} (Fig. 1).

After accounting for the protein, extra electron density on the concave surface remained, which corresponded to the H-trisaccharide antigen (fig. S2). Specificity between RBC36 and H-
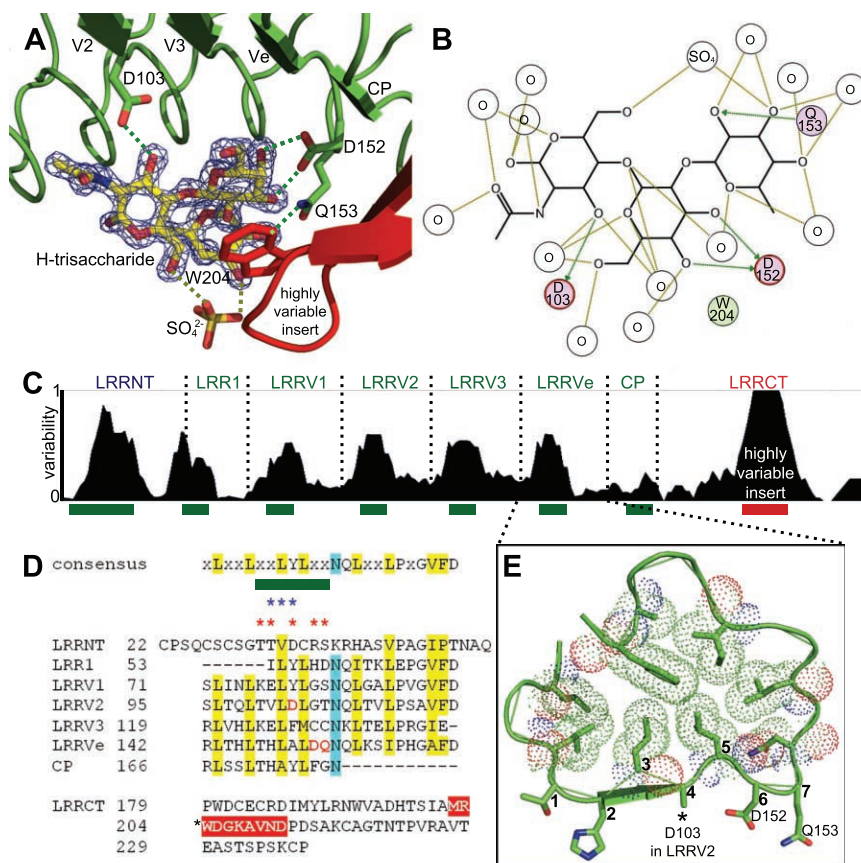
Fig. 2. The H-trisaccharide binding site of RBC36. (**A**) VLR residues involved in recognizing the H-trisaccharide. After refinement, a $2F_{obs} - F_{calc}$ electron density map was calculated and contoured at 2σ as a blue mesh around the H-trisaccharide. Colors are as in Fig. 1. (**B**) H-trisaccharide interaction with RBC36 including solvent molecules, modified from the ligand interaction calculation by the program MOE (33). O in a circle represents waters. Hydrogen bonds with RBC36 residues and solvent molecules are drawn with green and pale green lines, respectively. Trp^{204}, which is important for stabilizing the galactose via hydrophobic and stacking effects, is shown in a green circle beside the galactose sugar ring (14). (**C**) Sequence variability plot for amino acid residues from LRRNT to LRRCT of known VLRs. Green bars on the bottom represent residues on the concave surface; the red bar in the LRRCT shows the location of the highly variable insert. (**D**) Sequence alignment of LRR modules of RBC36 (14). The green bar shows the residues on the concave surface. Blue and red asterisks represent residues forming the β sheet and side chains that face the concave surface, respectively. Letters on yellow, blue, and red backgrounds show conserved hydrophobic residues, asparagine residues, and residues in the highly variable insert, respectively. Key residues on the concave surface (Asp^{103}, Asp^{152}, and Gln^{153}) for the H-trisaccharide interaction are shown as red letters; Trp^{204} in the highly variable insert is indicated by a black asterisk at the left. (**E**) The conformation of the LRRVe module highlights the tight packing of the conserved hydrophobic residues, with their van der Waals radii outlined in dots. The seven residues that form the concave surface are numbered from the N terminus to the C terminus of the LRR.
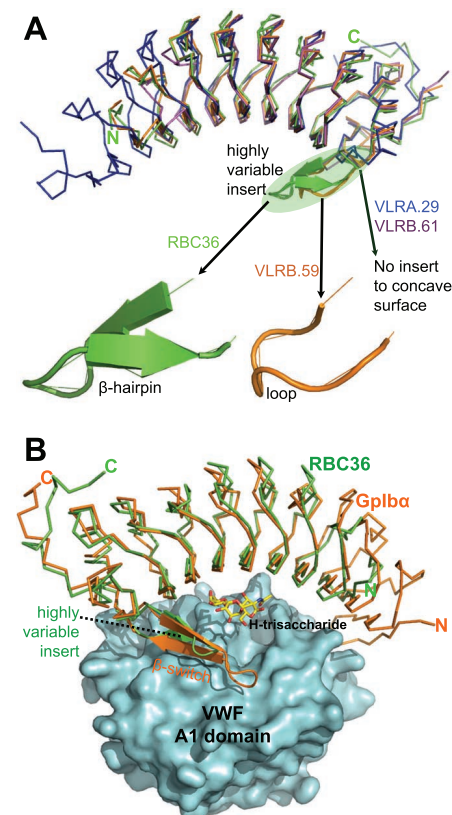
Fig. 3. Highly variable inserts of VLRs. (**A**) Crystal structures of lamprey RBC36 and three hagfish VLRs are superposed. Cα trace for different VLRs: RBC36 in green, VLRA.29 in blue (PDB ID 2O6Q), VLRB.59 in orange (PDB ID 2O6S), and VLRB.61 in magenta (PDB ID 2O6R), respectively. Highly variable inserts are drawn in cartoon representation. (**B**) Superposition of RBC36–H-trisaccharide complex and GpIbα-VWF A1 domain complex (PDB ID 1M10). Overall RBC36-ECD structure is rotated 180° vertically from (A) to highlight the comparison of the highly variable insert of RBC36 and the β switch of GpIbα. RBC36 is depicted as a green trace, GpIbα as an orange trace, VWF A1 domain as a surface representation in cyan, and the H-trisaccharide as in Fig. 1. The β hairpin of the highly variable insert of RBC36 and the β switch of GpIbα are shown in cartoon representation.

trisaccharide is mainly mediated by four hydrogen bonds (Fig. 2A) on the inner concave surface: between Asp^{103} O^{D2} and N-acetylglucosamine O^{AZ}, between Asp^{152} O^{D1} and galactose O$^{4'}$, between Asp^{152} O^{D2} and galactose O$^{3'}$, and between Gln^{153} N^{E2} and fucose O^{AL} (Fig. 2, A and B). Asp^{103} is located on LRRV2, and Asp^{152} and Gln^{153} on LRRVe.

With the concave surface of RBC36 firmly established as the antigen-binding site, we analyzed the variability in amino acids represented on this surface in other VLRs. From BLASTP searches (15) with RBC36, 24 VLR sequences were found with three LRRVs and sequence identity of >60%. The amino acid variation was higher on the concave surface of each LRR module (Fig. 2C, fig. S3, and table S2) and hence, to some extent, is analogous to the hypervariable regions (CDRs) in antibodies. In each canonical 24-residue LRR module, seven residues, xxL^{8}xL^{10}xx, are located on the concave surface and, of these, only the two obligate hydrophobic residues (L) face inward to form the hydrophobic core of the solenoidal structure (Fig. 2, D and E). Consequently, the other five residues could potentially contribute to antigen recognition, and correspond to the first, second, fourth, sixth, and seventh positions of this seven-residue segment in each LRR. Asp^{103}, Asp^{152}, and Gln^{153}, which contribute significantly to the interaction between RBC36 and H-trisaccharide, represent the fourth residue of the LRRV2 and the sixth

Fig. 4. The Interaction matrix of VLRs. (**A**) Seven residues on the concave surface of each LRR modules are shown as a main-chain stick model in the same view with Fig. 1C, and the Cα atoms are connected by black lines. Five residues of the seven on the concave face (1, 2, 4, 6, and 7) are available for the antigen recognition and are connected laterally by black lines; the third and fifth residues face inward to the hydrophobic core and are connected by yellow lines. (**B**) The simplified interaction matrix of (A). Residues involved in hydrogen bonds or in van der Waals contacts are labeled in a circle or in a square, respectively.

and seventh residues of the LRRVe concave surfaces. Eight other residues on the concave surface (His^{57}, Tyr^{79}, Thr^{106}, Phe^{127}, Cys^{129}, Ala^{150}, Tyr^{174}, and Phe^{176}) stabilize the H-trisaccharide interaction via 16 van der Waals contacts, as calculated with CONTACSYM (16). The carbohydrate antigen buries ~303 Å^2 on the VLR, whereas the corresponding buried surface on the antigen is ~246 Å^2 calculated with a 1.4 Å probe radius (17, 18), which is comparable to buried surfaces of haptens (~150 to 350 Å^2) with antibodies (19).

Another key interaction with the H-trisaccharide is between Trp^{204} and the galactose. The Trp^{204} indole is stacked parallel to the sugar ring, as observed in other sugar-protein complexes (Fig. 1 and Fig. 2A) (20). Trp^{204} is located in the middle of LRRCT, where the VLR sequences are extremely diverse and a highly variable insert is often present (21). The variability plot (Fig. 2C) illustrates that highly variable inserts of 2 to 12 residues occur in LRRCT (fig. S3). In RBC36, a 10-residue insert forms a β hairpin and Trp^{204} is located at the end of the first β strand, prior to the β-hairpin turn (Fig. 1 and Fig. 2A). Superposition of the crystal structures of lamprey RBC36 and the three hagfish VLRs reveals not only high sequence variability, but also secondary structure variation in their inserts. The VLRB.59 eight-residue insert is a loop, but similar in overall shape to the RBC36 β hairpin, whereas the VLRA.29 three-residue insert points toward the horseshoe side rather than its concave surface; VLRB.61 has no insert (Fig. 3A).

The RBC36 insert lies in close proximity to the concave β sheet, and its overall conformation is remarkably similar to the β switch in the C-terminal flank region of human glycoprotein Ibα (GpIbα) that interacts with von Willebrand factor (VWF) A1 domain (22) (Fig. 3B). A search for structural homologs of RBC36, using the DALI server (23), selected GpIbα (PDB ID 1M10) as one of the top three hits along with hagfish VLRB.59 (PDB ID 2O6S) and human Slit protein (PDB ID 2V9T), all of which are LRR-containing proteins. Interestingly, GpIbα and Slit also have crystal structures with their binding partners, in which their mode of interaction is very similar to the RBC36–H-trisaccharide complex where the same first, second, fourth, sixth, and seventh residues on the concave surface in each LRR module play the key role in ligand binding, but the insert in the Slit LRRCT does not contact its ligand (Fig. 3B and fig. S4). Considering that the secondary structures of the highly variable inserts of VLRs in the PDB are diverse (Fig. 3A), and the equivalent β switch of human GpIbα adopts a loop structure when GpIbα is crystallized by itself (22), these structure differences may also play an important role in antigen selection, recognition, and affinity. Conformational changes or isomerism in the inserts would also increase possible binding modes in much the same way as induced fit in the CDR loops of antibodies, especially for CDR H3 (24), or equivalent loops (α3, β3) in T cell receptors (25).

The concave molecular surface area of RBC36 is estimated to be ~1720 Å^2 calculated with a 1.4 Å probe radius (17, 18), compared with a buried surface of ~700 to 1000 Å^2 on average for the antigen-binding regions of immunoglobulins for proteins and other large antigens. Considering that the number of LRRVs in VLRs can be as many as seven and the concave surface area of one LRRV is about 220 Å^2, the total antigen-binding surface could extend to around 2600 Å^2 and could potentially accommodate binding sites for multiple antigens. However, this size could well be an overestimate as were the corresponding early predictions for the antigen-binding surfaces of antibodies (26), although the possibility of multiple paratopes in VLRs is certainly an intriguing concept.

With few exceptions, five residues on the concave surface of each LRR module are available for antigen binding (Fig. 2E and Fig. 4A). Hence, we mapped these key residues for antigen recognition onto a coordinate system that corresponds to the LRR modules along the x axis and key residue positions on the concave surface along the y axis (Fig. 4B). This interaction matrix should also be useful when other VLR-antigen complexes are determined.

Antigen recognition by antibodies in the vertebrate immune system is well documented and has revealed how the immunoglobulin (Ig) fold with its CDR loops can form a high-affinity binding site for virtually any antigen it encounters, whether natural or synthetic (24). Similarities and differences can now be assessed for antigen recognition by the Ig fold and the LRRs of VLRs, as well as TLRs. High sequence variability in the Ig fold is concentrated in CDRs H1, H2, H3, L1, L2, and L3, whereas that of VLRs is confined to the concave surface of each LRR module (Fig. 2C). In the Ig fold, a wide range of specificities and affinities for the antibody-combining sites is ensured not only by variability in amino acid composition, but also by insertions in the CDRs, especially in CDR H3 (27). However, VLRs contain few insertions on the concave surface of each LRR module, although diversity is attained by variation in the number (up to seven) and amino acid composition of the LRRV modules. The only insertion in VLRs is observed in the middle of LRRCT, which shows highly variable amino acid composition, length, and secondary structure. To what extent the highly variable insert in LRRCT contributes to the specificity, affinity, and shape of the antigen-binding site of VLRs awaits further VLR-antigen complex structures.

Recently, three crystal structures of TLR-ligand complexes—TLR4-MD2–Eritoran (28), TLR1-TLR2–lipopeptide (29), and TLR3-dsRNA (30)—have been determined. So far, only the binding mode of the TLR4-MD2 complex is similar to that of antigen recognition by VLRs, in that residues on the concave surface of the N-terminal and central domains of TLR4 interact with MD2. However, no interaction is seen between LRRCT of TLR4 and MD2 as observed between the highly variable insert in LRRCT of VLRs and

antigens. Because we do not yet have sufficient VLR and TLR complex structures to make statistically significant conclusions, and the number of LRR modules in TLRs is much greater than in VLRs, it may be too early to infer evolutionary relationships between VLRs and TLRs.

The crystal structure of RBC36-ECD in complex with the H-trisaccharide has provided structural insight into how VLRs recognize their antigens and provides a basis for rational design and modification of other antigen-specific VLRs. This VLR-antigen structure sheds light on the adaptation and evolution of primordial LRR proteins into their more specialized roles in pathogen recognition (e.g., TLRs) by the mammalian innate immune system.

References and Notes

1. Z. Pancer et al., Nature 430, 174 (2004).
2. M. N. Alder et al., Science 310, 1970 (2005).
3. When the distribution of LRRV modules per transcript was analyzed from 517 unique VLR sequences, VLRs were found to have an average of 1.31 LRRV modules between the canonical LRR1 and LRRVe modules (2): 109 VLRs (0 LRRV), 228 VLRs (1 LRRV), 119 VLRs (2 LRRVs), 45 VLRs (3 LRRVs), 6 VLRs (4 LRRVs), 8 VLRs (5 LRRVs), 1 VLR (6 LRRVs), and 1 VLR (7 LRRVs).
4. Z. Pancer et al., Proc. Natl. Acad. Sci. U.S.A. 102, 9224 (2005).
5. J. Choe, M. S. Kelker, I. A. Wilson, Science 309, 581 (2005); published online 16 June 2005 (10.1126/science.1115253).
6. J. K. Bell et al., Proc. Natl. Acad. Sci. U.S.A. 102, 10976 (2005).
7. H. M. Kim et al., J. Biol. Chem. 282, 6726 (2007).
8. B. R. Herrin et al., Proc. Natl. Acad. Sci. U.S.A. 105, 2040 (2008).
9. M. N. Alder et al., Nat. Immunol. 9, 319 (2008).
10. G. A. Boffa, J. M. Fine, A. Drilhon, P. Amouch, Nature 214, 700 (1967).
11. B. Pollara, G. W. Litman, J. Finstad, J. Howell, R. A. Good, J. Immunol. 105, 738 (1970).
12. At least four subtypes of H-antigens are known, which can be further converted into the A-antigen of human blood group A and B-antigen of human blood group B by glycosyltransferases A and B, respectively (31) (fig. S1).
13. See supporting material on Science Online.
14. Single-letter abbreviations for amino acid residues: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.
15. S. F. Altschul et al., Nucleic Acids Res. 25, 3389 (1997).
16. S. Sheriff, W. A. Hendrickson, J. L. Smith, J. Mol. Biol. 197, 273 (1987).
17. B. R. Gelin, M. Karplus, Biochemistry 18, 1256 (1979).
18. M. L. Connolly, J. Mol. Graph. 11, 139 (1993).
19. I. A. Wilson, R. L. Stanfield, Curr. Opin. Struct. Biol. 3, 113 (1993).
20. Aromatic residues, such as Trp and Tyr, are frequently observed in the protein sugar-binding sites. For example, in the maltose/maltooligosaccharide-binding protein (32), aromatic residues play a major role in carbohydrate recognition and binding by providing not only hydrophobic stacking interactions between aromatic residues and the hydrophobic face of the sugar rings, but also polar contacts to the sugar hydroxyl groups.
21. I. B. Rogozin et al., Nat. Immunol. 8, 647 (2007).
22. E. G. Huizinga et al., Science 297, 1176 (2002).
23. L. Holm, C. Sander, J. Mol. Biol. 233, 123 (1993).
24. I. A. Wilson, R. L. Stanfield, Curr. Opin. Struct. Biol. 4, 857 (1994).
25. K. C. Garcia, L. Teyton, I. A. Wilson, Annu. Rev. Immunol. 17, 369 (1999).
26. P. von Gara, Z. Immunitaetsforsch. Exp. Ther. 71, 1 (1931).
27. T. T. Wu, G. Johnson, E. A. Kabat, Proteins 16, 1 (1993).
28. H. M. Kim et al., Cell 130, 906 (2007).
29. M. S. Jin et al., Cell 130, 1071 (2007).
30. L. Liu et al., Science 320, 379 (2008).
31. F. Yamamoto, P. D. McNeill, S. Hakomori, Biochem. Biophys. Res. Commun. 187, 366 (1992).
32. X. Duan, F. A. Quiocho, Biochemistry 41, 706 (2002).
33. Molecular Operating Environment (MOE), version 2000.09 (Chemical Computing Group, Montreal, 2003).
34. We thank R. L. Stanfield, X. Dai, and X. Zhu for help with data collection and analysis; M. N. Alder for help with VLR cDNA library preparation; J. Paulson and R. McBride for helpful comments and suggestions; and J. Vanhnasy, B. Droese, and H.-J. Kim for technical support and advice. Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. Supported by NIH grants AI42266 (I.A.W.) and AI072435 (M.D.C.), the Georgia Research Alliance (M.D.C.), and the Skaggs Institute for Chemical Biology (I.A.W.). This is Scripps Research Institute manuscript 19581-MB. Coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession code 3E6J.

Disruption of the *CFTR* Gene Produces a Model of Cystic Fibrosis in Newborn Pigs

Christopher S. Rogers,[1]* David A. Stoltz,[1]* David K. Meyerholz,[2]* Lynda S. Ostedgaard,[1] Tatiana Rokhlina,[1] Peter J. Taft,[1] Mark P. Rogan,[1] Alejandro A. Pezzulo,[1] Philip H. Karp,[1,3] Omar A. Itani,[1] Amanda C. Kabel,[1] Christine L. Wohlford-Lenane,[4] Greg J. Davis,[1] Robert A. Hanfland,[5] Tony L. Smith,[5] Melissa Samuel,[6] David Wax,[6] Clifton N. Murphy,[6] August Rieke,[6] Kristin Whitworth,[6] Aliye Uc,[4] Timothy D. Starner,[4] Kim A. Brogden,[7] Joel Shilyansky,[5] Paul B. McCray Jr.,[4] Joseph Zabner,[1] Randall S. Prather,[6] Michael J. Welsh[1,3,8]†

Almost two decades after *CFTR* was identified as the gene responsible for cystic fibrosis (CF), we still lack answers to many questions about the pathogenesis of the disease, and it remains incurable. Mice with a disrupted *CFTR* gene have greatly facilitated CF studies, but the mutant mice do not develop the characteristic manifestations of human CF, including abnormalities of the pancreas, lung, intestine, liver, and other organs. Because pigs share many anatomical and physiological features with humans, we generated pigs with a targeted disruption of both *CFTR* alleles. Newborn pigs lacking CFTR exhibited defective chloride transport and developed meconium ileus, exocrine pancreatic destruction, and focal biliary cirrhosis, replicating abnormalities seen in newborn humans with CF. The pig model may provide opportunities to address persistent questions about CF pathogenesis and accelerate discovery of strategies for prevention and treatment.

U nderstanding human disease often requires the use of animal models. Mice have been the overwhelming species of choice because methods for specifically altering their genome have been readily available. However, in many cases, mice with targeted gene manipulations fail to replicate phenotypes observed in humans—one example is cystic fibrosis (CF).

In 1938, Dorothy Andersen coined the term "cystic fibrosis of the pancreas" (1). Over the ensuing years, investigators learned that CF involved many other organs, including the intestine, lung, sweat gland, liver, gallbladder, and male genital tract (2–4). We now know CF to be a common, autosomal recessive disease with a carrier rate of ~5% in Caucasians. In 1989, the gene mutated in CF was identified, and its product was named cystic fibrosis transmembrane conductance regulator (CFTR) (5). Soon thereafter, it was discovered that CFTR is a regulated anion channel that may also affect other transport processes (3).

Despite many advances, our understanding of CF pathogenesis remains incomplete, thus hindering the development of new therapies. This begs the question: Why have we not made more progress? Whereas superb clinical research has guided thoughts about CF, interpretations about pathogenesis are often based on observations ob-

[1]Department of Internal Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA. [2]Department of Pathology, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA. [3]Howard Hughes Medical Institute (HHMI), Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA. [4]Department of Pediatrics, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA. [5]Department of Surgery, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA. [6]Division of Animal Sciences, University of Missouri, Columbia, MO 65211, USA. [7]Department of Periodontics and Dows Institute for Dental Research, College of Dentistry, University of Iowa, Iowa City, IA 52242, USA. [8]Department of Molecular Physiology and Biophysics, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: michael-welsh@uiowa.edu

tained long after the disease onset, and many studies cannot be carried out in humans. Cell-based models have also proven valuable for research (6) but are limited because the disease involves a whole organism. Gene-targeted mouse models have likewise been instructive about CF, yet the mutant mice do not develop the pancreatic, airway, intestinal, or liver disease typically found in humans (7–9).

To develop a new CF model, we chose pigs because in terms of anatomy, biochemistry, physiology, size, life span, and genetics, they are more similar to humans than are mice (10, 11). We used homologous recombination in fibroblasts of outbred domestic pigs to disrupt the *CFTR* gene and somatic cell nuclear transfer to generate *CFTR*$^{+/-}$ pigs (12).

At sexual maturity (~6 to 7 months), female *CFTR*$^{+/-}$ pigs were bred to *CFTR*$^{+/-}$ males. Six litters produced 64 piglets. Genotyping (Fig. 1A) revealed 18 *CFTR*$^{+/+}$, 26 *CFTR*$^{+/-}$, and 20 *CFTR*$^{-/-}$ animals, a ratio not significantly different from the expected ratio of 1:2:1 (13). Figure 1B shows the first litter. Birth weights varied but did not segregate by genotype (Fig. 1C). The piglets looked normal at birth, and their genotype could not be discerned by appearance. A normal appearance is consistent with findings in humans.

Northern blot analysis and quantitative reverse transcription polymerase chain reaction (RT-PCR) did not detect normal *CFTR* transcripts (Fig. 1A). Immunoprecipitation detected no normal CFTR protein. Like human CFTR (14, 15), in wild-type (WT) tissue the

porcine protein localized apically in airway epithelia and ileal crypts (Fig. 1D).

We assessed CFTR function in vivo by measuring transepithelial voltage (Vt) across nasal epithelia (16) (Fig. 1, E and F). As in humans with CF, baseline Vt was hyperpolarized in *CFTR*$^{-/-}$ piglets. Amiloride, which inhibits ENaC Na$^+$ channels, reduced Vt in all genotypes. To test for CFTR channel activity, we perfused the apical surface with a Cl$^-$-free solution and added isoproterenol to increase cellular levels of cyclic adenosine monophosphate; these interventions hyperpolarized nasal Vt in WT and heterozygous pigs, but not *CFTR*$^{-/-}$ animals. Perfusion with adenosine triphosphate to activate P2Y2 receptors and Ca^{2+}-activated Cl$^-$ channels (17) further hyperpolarized Vt, and the response did not differ significantly

Fig. 1. *CFTR*$^{-/-}$ piglets appear normal at birth. (**A**) Upper panel depicts insertion into porcine *CFTR* exon 10 of a phosphoglycerate kinase (PGK) promoter (yellow) driving a neomycin resistance cDNA (orange), and an engineered stop codon. Position of probe (green), PCR primers (arrowheads), and *Bgl*II sites (B) is indicated. The second and third panels show genotyping by PCR and Southern blot analysis of genomic DNA. Lanes C1, C2, and C3 contain controls of *CFTR*$^{+/+}$, $^{+/-}$, and $^{-/-}$ DNA, respectively. The fourth panel shows Northern blot analysis of ileal *CFTR* and *GAPDH* mRNA. Consistent with the Northern blot, quantitative RT-PCR of exon 10 (the targeted site) detected <0.1% of *CFTR* transcripts in CFTR$^{-/-}$ ileum, relative to *CFTR*$^{+/+}$ (n = 6 and 4 piglets, respectively). The bottom panel shows immunoprecipitation (IP) and phosphorylation of CFTR plus recombinant CFTR in baby hamster kidney cells. (**B**) First litter containing piglets of all three genotypes. (**C**) Birth weights. Mean ± SD of weights: 1.31 ± 0.24 kg for *CFTR*$^{+/+}$ (orange), 1.35 ± 0.28 kg for *CFTR*$^{+/-}$ (green), and 1.31 ± 0.23 kg for *CFTR*$^{-/-}$ (blue) animals. (**D**) Immunocytochemistry of CFTR in airway epithelia (top) and ileum (bottom). Figures are differential interference contrast with staining for ZO-1 (a component of tight junctions, red), CFTR (green), and nuclei (4′,6′-diamidino-2-phenylindole, blue). See also fig. S1. Scale bars, 10 μm. (**E**) Tracings of in vivo nasal Vt measured in newborn pigs. After baseline measurements, the following agents/solutions were sequentially added to the epithelial perfusate: amiloride (100 μM), Cl$^-$-free solution, isoproterenol (10 μM), ATP (100 μM), and GlyH-101 (100 μM). (**F**) Average nasal Vt measurements as indicated in (E). Data from four *CFTR*$^{+/+}$ and four *CFTR*$^{+/-}$ piglets (gray squares) were not statistically different and were combined and compared with data from five *CFTR*$^{-/-}$ piglets (blue circles). Values of baseline nasal Vt for *CFTR*$^{-/-}$ piglets differed from the controls, as did the changes in Vt induced by adding amiloride, a Cl$^-$-free solution, and GlyH-101 (all P < 0.05). Data are mean ± SEM.
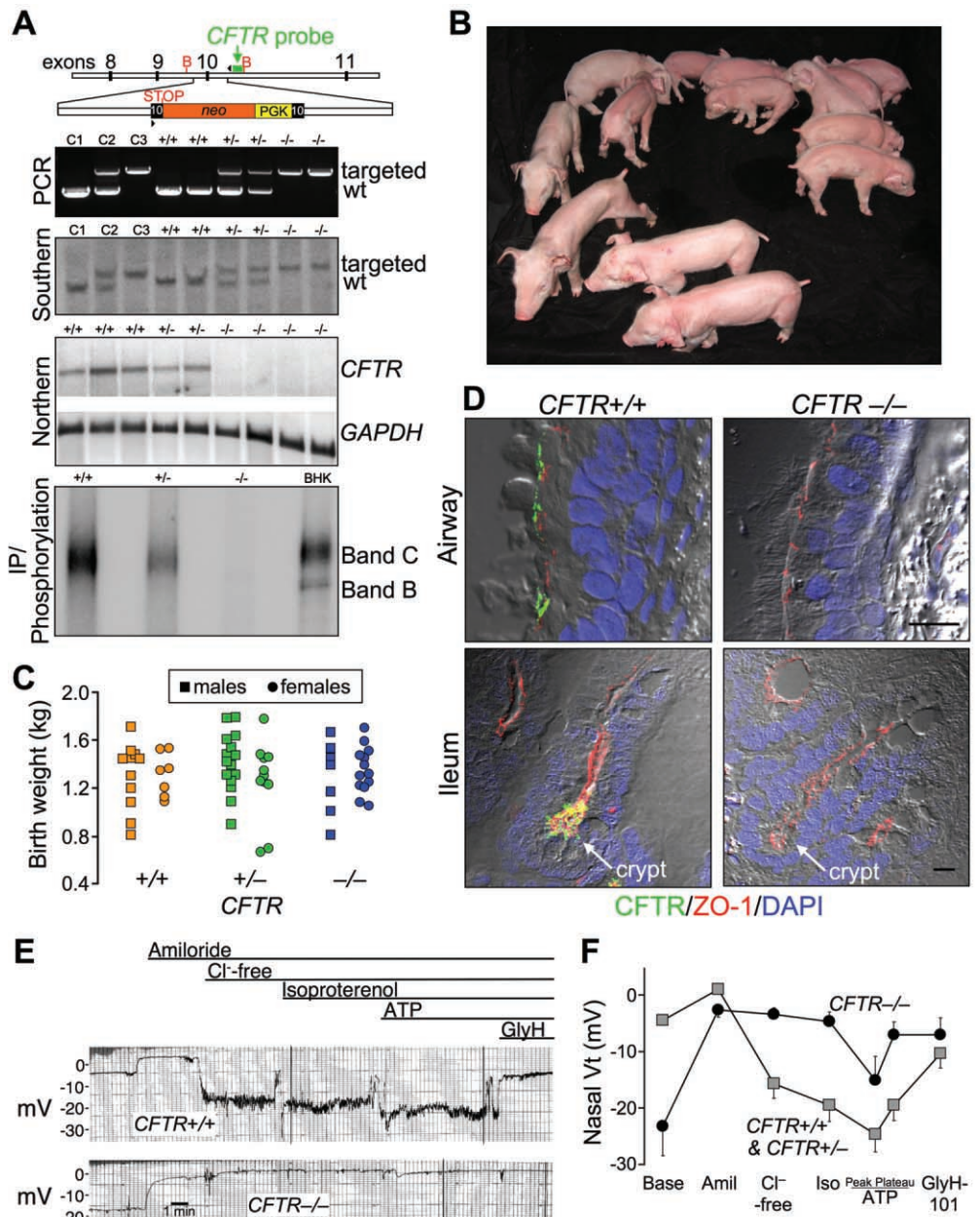
between genotypes. Perfusion with the CFTR inhibitor GlyH-101 (*18*) depolarized Vt in controls animals, but not in $CFTR^{-/-}$ piglets. These data reveal the loss of CFTR Cl⁻ channel activity in newborn $CFTR^{-/-}$ pigs. Whereas the lack of data from newborn humans precludes a direct comparison, our data qualitatively match those from adults and children with CF (*16*).

What phenotypes would be expected if newborn $CFTR^{-/-}$ piglets model human disease? Figure 2A shows some human CF phenotypes and the time spans when they become clinically apparent (*3*, *19*). The earliest manifestation (hours to 2 days) is meconium ileus, an intestinal obstruction occurring in ~15% of CF infants (*2*, *3*, *19*, *20*). Obstruction can occur throughout the small intestine or colon, but it is most often observed near the ileocecal junction. Distal to the obstruction, the bowel is small and atretic (microcolon). Intestinal perforation in utero or postnatally occurs in some infants.

Newborn $CFTR^{-/-}$ piglets failed to pass feces or gain weight (Fig. 2B). By 24 to 40 hours of age, they stopped eating, developed abdominal distension, and had bile-stained emesis, which are clinical signs of intestinal obstruction. We examined histopathology between birth and 12 hours in piglets that had not eaten and between 24 and 40 hours in piglets that were fed colostrum and milk replacer. Except as noted, the pathologic changes refer to the earlier time period. After 30 to 40 hours, the stomachs of $CFTR^{-/-}$ piglets contained small amounts of green, bile-stained milk (Fig. 2C). The proximal small intestine was dilated by small amounts of milk and abundant gas. The site of obstruction ranged from mid-distal small intestine to proximal spiral colon, the anatomical equivalent of the human ascending colon. Perforation and peritonitis occurred in some piglets. Dark green meconium distended the $CFTR^{-/-}$ intestine, and adjacent villi showed degeneration and atrophy, whereas $CFTR^{+/+}$ ileum had long villi

(Fig. 2D). Distal to the meconium, luminal diameter was reduced with mild-to-severe mucinous hyperplasia, including mucoid luminal "plugs" (Fig. 2E). These changes replicate those seen in humans with CF (*3*, *19*).

The penetrance of meconium ileus was 100% in $CFTR^{-/-}$ piglets versus ~15% in human CF. Potential explanations for this difference include a restricted genetic background in our pig model versus that in humans, a null mutation in the pigs versus mutations in humans that might yield tiny amounts of protein function, and anatomical or physiological differences [see footnote F1 in (*13*)].

Exocrine pancreatic insufficiency afflicts 90 to 95% of patients with CF (*1–4*, *19*, *21*, *22*). The porcine $CFTR^{-/-}$ pancreas was small (Fig. 3A), and microscopic examination revealed small, degenerative lobules with increased loose adipose and myxomatous tissue, as well as scattered-to-moderate cellular inflammation (Fig. 3, B and C). Residual acini had diminished amounts of eosinophilic zymogen granules (Fig. 3D). Centroacinar spaces, ductules, and ducts were variably dilated and obstructed by eosinophilic material plus infrequent neutrophils and macrophages mixed with cellular debris (Fig. 3E). Ducts and ductules had foci of mucinous metaplasia. Pancreatic endocrine tissue was spared (Fig. 3C). These changes are similar to those originally described by Andersen and others (*1*, *19*, *21*) in their studies of human CF.

Humans often require surgery to relieve meconium ileus (*3*). We performed an ileostomy on one $CFTR^{+/+}$ piglet and three $CFTR^{-/-}$ piglets. Because of technical problems in postoperative care, only one piglet (genotype $CFTR^{-/-}$) out of the four recovered. Once newborn infants recover from meconium ileus, the course of their disease resembles that in other patients. Likewise, the piglet grew and went on to develop pancreatic insufficiency. At 10 weeks, the piglet had an episode resembling "distal intestinal obstruction syndrome," which was successfully treated as in humans (*3*) [also see footnote F2 in (*13*)]. Although these observations were based on a single animal, they bore a marked resemblance to clinical observations of human CF.

Focal biliary cirrhosis is the second most common cause of CF mortality (*3*, *20*, *23*). The porcine $CFTR^{-/-}$ liver revealed infrequent, mild-to-moderate hepatic lesions (Fig. 3F). Chronic cellular inflammation, ductular hyperplasia, and mild fibrosis were typical of focal biliary cirrhosis. Gallbladder abnormalities, including gallstones, occur in 15 to 30% of patients, and a small gallbladder is a common autopsy finding (*3*). Similarly, porcine $CFTR^{-/-}$ gallbladders were small and often filled with congealed bile and mucus (Fig. 3, G and H). Epithelia showed diffuse mucinous changes with folds extending into the lumen.

Approximately 97% of males with CF are infertile (*3*); the vas deferens is often normal at birth, and obstruction is thought to cause progressive deterioration [see footnote F3 in (*13*)]. In all piglets, the vas deferens appeared to be intact. Paranasal sinus abnormalities occur in most chil-
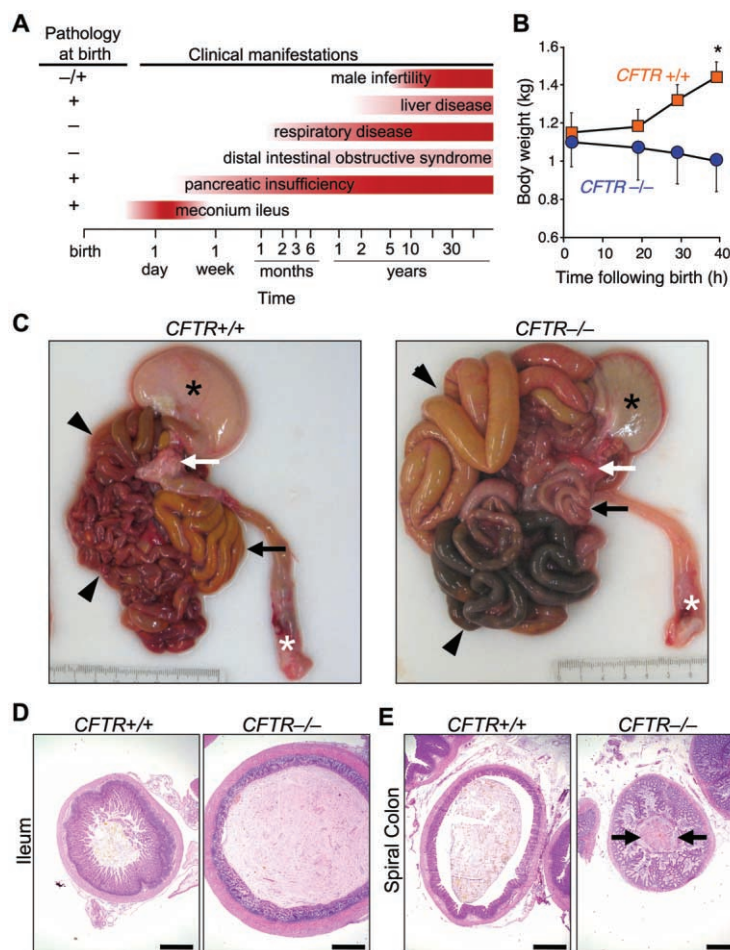
Fig. 2. $CFTR^{-/-}$ piglets develop meconium ileus. (**A**) Schematic shows some clinical and histopathological CF manifestations. Pathological abnormalities are present before clinical disease becomes apparent. (**B**) Weight after birth. Animals were fed colostrum and milk-replacer. $n = 7$ $CFTR^{+/+}$ and 4 $CFTR^{-/-}$ piglets. Data are mean ± SEM. *$P < 0.05$. (**C**) Gross appearance of gastrointestinal tract. Piglets were fed colostrum and milk-replacer for 30 to 40 hours and then euthanized. Stomach, black asterisk; small intestine, arrowheads; pancreas, white arrow; rectum, white asterisk; and spiral colon, black arrow. Of 16 $CFTR^{-/-}$ piglets, the obstruction occurred in the small intestine in 7 animals and in spiral colon in 9. (**D** and **E**) Microscopic appearance of the ileum (D) and colon (E). Hematoxylin and eosin (H&E) stain. Scale bars, 1 mm. Images are representative of severe meconium ileus occurring in 16 out of 16 (16/16) $CFTR^{-/-}$ piglets.

Fig. 3. $CFTR^{-/-}$ piglets have exocrine pancreatic destruction and liver and gallbladder abnormalities. (**A**) Gross appearance of pancreas. Scale bar, 0.5 cm. (**B**) Loss of parenchyma in the $CFTR^{-/-}$ pancreas. H&E stain. Scale bars, 500 μm. (**C**) Pancreatic ducts and islets of Langerhans (arrowheads). Scale bars, 100 μm. (**D**) $CFTR^{-/-}$ ductules and acini dilated by eosinophilic inspissated material that formed concentrically lamellar concretions (arrows and inset). H&E stain. Scale bars, 33 μm. (**E**) Ducts within the $CFTR^{-/-}$ pancreas. H&E stain, left; periodic acid–Schiff (PAS) stain, right. Scale bars, 50 μm. (**F**) Microscopic appearance of liver. H&E stain. Arrows indicate focal expansion of portal areas by chronic cellular inflammation. Scale bars, 100 μm. (**G**) Gross appearance of gallbladder. When the $CFTR^{+/+}$ gallbladder was sectioned, bile drained away rapidly with collapse of the mucosal wall. $CFTR^{-/-}$ bile was congealed (arrow) and retained in the lumen of a smaller gallbladder. Scale bar, 0.5 cm. (**H**) Microscopic appearance of gallbladder. $CFTR^{-/-}$ gallbladders had congealed, inspissated bile (asterisk) with variable mucus production (arrows, H&E stain) highlighted as a magenta color in PAS stained tissue. Scale bars, 500 μm. Images are representative of severe pancreatic lesions (15/15 $CFTR^{-/-}$ piglets), mild-to-moderate liver lesions (3/15), and mild-to-severe gallbladder/duct lesions (15/15).
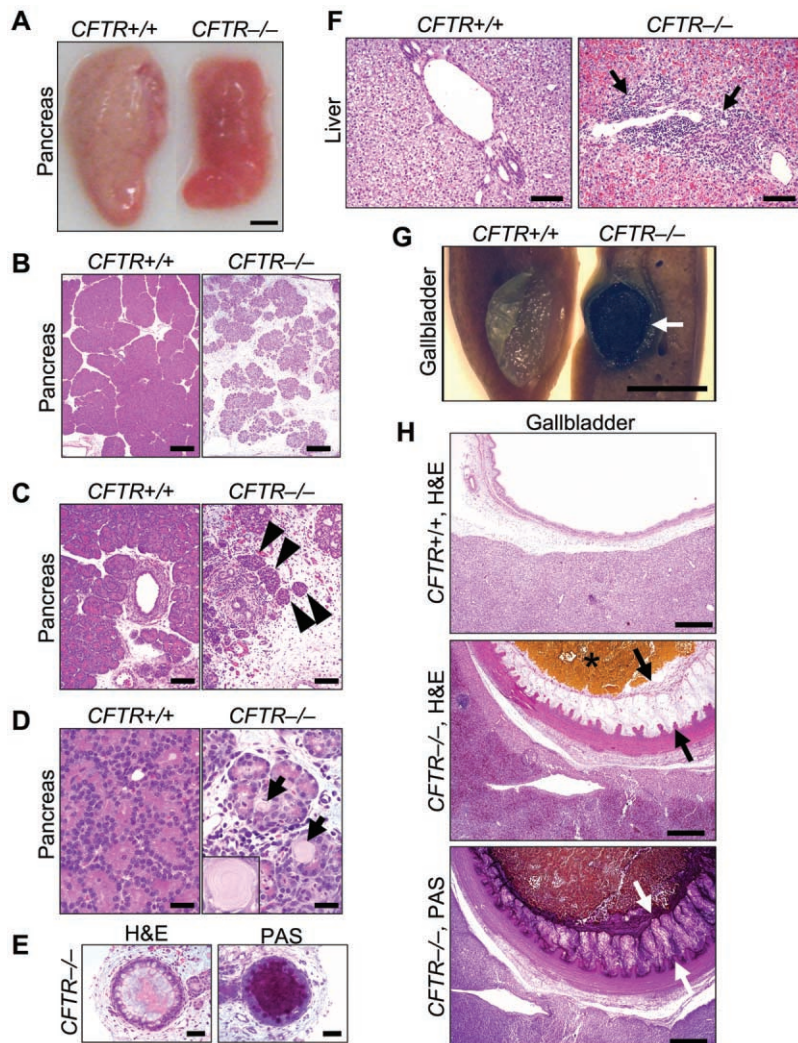
Fig. 4. The lungs of newborn $CFTR^{-/-}$ and $CFTR^{+/+}$ piglets appear normal. (**A**) Microscopic appearance of lung from piglets <12 hours old. H&E staining. Scale bars, 1 mm (left) and 50 μm (right). (**B**) Bronchial epithelia and submucosal glands. H&E staining. Scale bars, 50 μm. Images are representative of the lack of lesions in 15 out of 15 $CFTR^{-/-}$.
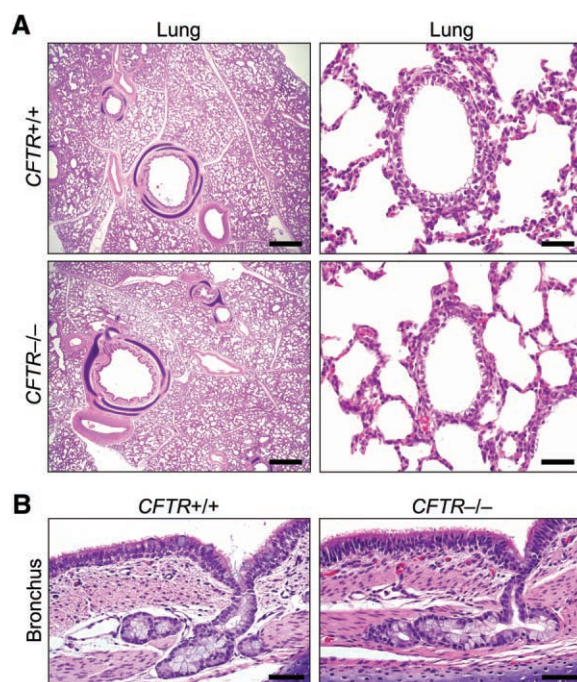
dren and adults with CF (3). Although $CFTR^{-/-}$ porcine paranasal sinuses showed no abnormalities, this negative result is difficult to interpret because it is unclear when sinus disease develops in humans. The salivary glands, nasal cavity, esophageal glands, kidney, heart, striated muscle, spleen, adrenals, eyes, brain, skin, and eccrine sweat glands on the snout revealed no abnormalities in $CFTR^{-/-}$ piglets. In all tissues, we observed no differences between WT and heterozygous $CFTR^{+/-}$ animals.

Lung disease is the major cause of CF morbidity and mortality (2–4). The onset of clinical respiratory manifestations varies, with some patients developing symptoms a few months after birth and others after several years. Eventually, most patients develop chronic airway infection and inflammation that destroy the lung. The lungs of neonatal $CFTR^{-/-}$ piglets appeared similar to those of their WT littermates. $CFTR^{-/-}$ lungs lacked evidence of cellular inflammation in airways or parenchyma (Fig. 4A). Airway epithelia and submucosal glands appeared similar in all three genotypes, and we found no evidence of dilated or plugged submucosal gland ducts [see footnote F4 in (13)] (Fig. 4B). Bronchoalveolar lavage 6 to

12 hours after birth showed no evidence of infection, and there were no significant differences between cell counts or levels of interleukin-8 (IL-8) across genotypes (figs. S2 and S3) (*13*).

Whether airway inflammation precedes infection in CF patients or vice versa has been a persistent question. Studies of bronchoalveolar lavage in infants and young children have both supported and argued against the presence of inflammation (increased IL-8 and neutrophilia) without infection (*24*, *25*). In vitro airway epithelial models have also given conflicting results (*26*, *27*). Studies of human fetal trachea transplanted into mice suggested inflammation might occur in developing CF airways (*28*). Although our data do not resolve this controversy, we had the advantage of studying lungs between birth and 12 hours of age, and we found no evidence of abnormal infection or inflammation. Tracking the lungs as $CFTR^{-/-}$ piglets are exposed to additional environmental challenges may inform our understanding of how respiratory disease develops in children and young adults.

The clinical, electrophysiological, and pathological findings in newborn $CFTR^{-/-}$ pigs were markedly similar to those in human neonates with CF (table S1) (*13*). Abdominal lesions dominate the initial presentation in both, with identical appearance of meconium ileus and exocrine pancreatic destruction. In addition, as in humans, the piglets have hepatic changes consistent with early focal biliary cirrhosis and abnormalities of the gallbladder and bile ducts. The lack of abnormalities in the vas deferens and lungs at birth is another similarity. Overall, these encouraging results suggest that the pig model may provide investigators with further opportunities to study CF and develop strategies for prevention and treatment.

References and Notes

1. D. H. Andersen, *Am. J. Dis. Child.* **56**, 344 (1938).
2. P. M. Quinton, *Physiol. Rev.* **79**, S3 (1999).
3. M. J. Welsh, B. W. Ramsey, F. Accurso, G. R. Cutting, in *The Metabolic and Molecular Basis of Inherited Disease*, C. R. Scriver et al., Eds. (McGraw-Hill, New York, 2001), pp. 5121–5189.
4. S. M. Rowe, S. Miller, E. J. Sorscher, *N. Engl. J. Med.* **352**, 1992 (2005).
5. J. R. Riordan et al., *Science* **245**, 1066 (1989).
6. P. H. Karp et al., in vol. 188 of *Epithelial Cell Culture Protocols*, C. Wise, Ed. (Humana, Totowa, NJ, 2002), pp. 115–137.
7. B. R. Grubb, R. C. Boucher, *Physiol. Rev.* **79**, S193 (1999).
8. C. Guilbault, Z. Saeed, G. P. Downey, D. Radzioch, *Am. J. Respir. Cell Mol. Biol.* **36**, 1 (2007).
9. L. L. Clarke, L. R. Gawenis, C. L. Franklin, M. C. Harline, *Lab. Anim. Sci.* **46**, 612 (1996).
10. Z. Ibrahim et al., *Xenotransplantation* **13**, 488 (2006).
11. C. S. Rogers et al., *Am. J. Physiol. Lung Cell Mol. Physiol.* **295**, L240 (2008).
12. C. S. Rogers et al., *J. Clin. Invest.* **118**, 1571 (2008).
13. Supporting material is available on *Science* Online.
14. I. Crawford et al., *Proc. Natl. Acad. Sci. U.S.A.* **88**, 9262 (1991).
15. G. M. Denning, L. S. Ostedgaard, S. H. Cheng, A. E. Smith, M. J. Welsh, *J. Clin. Invest.* **89**, 339 (1992).
16. T. A. Standaert et al., *Pediatr. Pulmonol.* **37**, 385 (2004).
17. S. J. Mason, A. M. Paradiso, R. C. Boucher, *Br. J. Pharmacol.* **103**, 1649 (1991).
18. C. Muanprasat et al., *J. Gen. Physiol.* **124**, 125 (2004).
19. E. H. Oppenheimer, J. R. Esterly, *Perspect. Pediatr. Pathol.* **2**, 241 (1975).
20. M. Wilschanski, P. R. Durie, *J. R. Soc. Med.* **91**, 40 (1998).
21. J. R. Imrie, D. G. Fagan, J. M. Sturgess, *Am. J. Pathol.* **95**, 697 (1979).
22. S. M. Blackman et al., *Gastroenterology* **131**, 1030 (2006).
23. E. H. Oppenheimer, J. R. Esterly, *J. Pediatr.* **86**, 683 (1975).
24. T. Z. Khan et al., *Am. J. Respir. Crit. Care Med.* **151**, 1075 (1995).
25. D. S. Armstrong et al., *Pediatr. Pulmonol.* **40**, 500 (2005).
26. A. A. Stecenko et al., *Inflammation* **25**, 145 (2001).
27. N. Aldallal et al., *Am. J. Respir. Crit. Care Med.* **166**, 1248 (2002).
28. R. Tirouvanziam et al., *Am. J. Respir. Cell Mol. Biol.* **23**, 121 (2000).
29. We thank A. Arias, E. Bagnall, J. Bartlett, B. Bauer, T. Bohnert, E. Burnight, K. Dobbs, C. Dohrn, L. Dowell, L. Gakhar, D. Guilliams, Y. Hao, S. C. Isom, S. Jones, S. Korte, B. Kussman, J. Launspach, M. Linville, E. Mahan, T. Mayhew, C. McHughes, K. Munson, M. Parker, C. Randak, S. Ramachandran, J. Ross, A. Small, L. Spate, D. Vermeer, E. Walters, and J. Whyte for excellent assistance. We thank our patients with CF for inspiration, and we also thank Cystic Fibrosis Foundation Therapeutics and R. Bridges for the gift of GlyH-101. This work was supported by the National Heart Lung and Blood Institute (grant HL51670), the National Institute of Diabetes and Digestive and Kidney Diseases (grant DK54759), Food for the 21st Century, the Cystic Fibrosis Foundation, and HHMI. C.S.R. was supported by NIH training grant HL07638. D.A.S. is a Parker B. Francis Fellow and was supported by the National Institute of Allergy and Infectious Diseases (grant AI076671). M.J.W. is an Investigator of the HHMI. C.S.R., R.S.P., M.J.W., and the University of Iowa Foundation have applied for a patent related to the work reported in this paper, and C.S.R. and M.J.W. are founders of Exemplar Genetics, a company that is licensing materials and technology related to this work. C.S.R. is currently Director of Research and Development at Exemplar Genetics.

Seeding and Propagation of Untransformed Mouse Mammary Cells in the Lung

Katrina Podsypanina,* Yi-Chieh Nancy Du, Martin Jechlinger, Levi J. Beverly, Dolores Hambardzumyan, Harold Varmus

The acquisition of metastatic ability by tumor cells is considered a late event in the evolution of malignant tumors. We report that untransformed mouse mammary cells that have been engineered to express the inducible oncogenic transgenes MYC and Kras^{D12}, or polyoma middle T, and introduced into the systemic circulation of a mouse can bypass transformation at the primary site and develop into metastatic pulmonary lesions upon immediate or delayed oncogene induction. Therefore, previously untransformed mammary cells may establish residence in the lung once they have entered the bloodstream and may assume malignant growth upon oncogene activation. Mammary cells lacking oncogenic transgenes displayed a similar capacity for long-term residence in the lungs but did not form ectopic tumors.

Metastatic dissemination of cancer cells, the major cause of cancer mortality, is traditionally viewed as a late-stage event (*1*), although mammary epithelial cells have been shown to disseminate systemically from early neoplastic lesions in transgenic mice and from ductal carcinoma in situ in women (*2*). There is ample evidence that the ability of fully transformed tumor cells to metastasize depends on the regulation of developmental programs and external environmental cues (*3–11*), but to what extent the seeding or growth of tumor cells at the ectopic site is dependent on the initiating transforming event(s) is a subject of debate (*12*). We have developed a system that separates the process of seeding cells in the lung from the process of malignant growth at an ectopic site by using animals engineered to express potent oncogenes in a doxycycline-dependent mammary-specific manner. After intravenous (IV) injection of marked mammary cells that have different genetic potentials [no oncogenes, or oncogenes that will be expressed only after cells have taken up residence in an ectopic site (the lungs)], normal mammary cells can lodge in the lungs, grow slowly, and become frank metastatic malignancies once potent oncogenes are turned on.

We recently described tri-transgenic *TetO-MYC;TetO-Kras^{D12};MMTV-rtTA* (*TOM;TOR; MTB*) mice that coordinately express *MYC* and mutant *Kras* oncogenes in mammary epithelial cells when fed doxycycline (*13*). Doxycycline-naïve animals do not express the transgenic oncogenes and have morphologically and functionally normal mammary glands, but they develop diffuse autochthonous tumors within 3 to 4 weeks after doxycycline exposure. Tumors that form because of the expression of these oncogenes display malignant characteristics, such as transplantability and metastasis (fig. S1). Therefore, this model provides primary mammary cells that can be switched from a normal to a neoplastic state with simple experimental manipulation.

Program in Cancer Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA.

*To whom correspondence should be addressed. E-mail: podsypak@mskcc.org

Fig. 1. Untransformed mouse mammary cells form lung metastases after IV injection and induction of oncogenes. (**A** to **D**) Lung metastases develop from intravenously injected phenotypically normal mammary cells upon activation of *MYC* and *Kras*[D12] transgenes. (**A**) Metastases were monitored by MRI in *Rag1*[−/−] mice after IV delivery of 1×10^6 dissociated primary mammary cells from doxycycline-naïve *TOM;TOR;MTB* mice. Recipient mice were fed doxycycline for 6 weeks starting 1 day before injection. Representative axial (top) and coronal (bottom) images obtained from the same animal 4 and 6 weeks after injection show development of a solid nodule (arrowheads) in the lung. (**B**) Foci of hematoxylin-eosin (H/E)–stained mammary adenocarcinoma (arrowheads) in paraffin-embedded lung sections of the same *Rag1*[−/−] mouse as in (A). Scale bars indicate 1 mm (left) and 0.1 mm (right). (**C**) No tumors were observed in lung sections of *Rag1*[−/−] mice that did not receive doxycycline after IV delivery of 1×10^6 primary mammary cells from doxycycline-naïve *TOM;TOR;MTB* mice. Scale bar, 1 mm. (**D**) Tumor cells from the same animal as in (A), but not the surrounding lung tissue, stained with anti-MYC antisera. Scale bar, 0.1 mm. (**E** and **F**) Lung metastases develop from intravenously injected phenotypically normal mammary cells upon activation of a *PyMT* transgene. (**E**) Donor cells expressing their transgene were detected by biolumi-

nescence imaging after 5×10^5 primary mammary cells from doxycycline-naïve *TOMT:IRES:Luc;MTB* mice were injected intravenously into *Rag1*[−/−] mice that were placed on doxycycline 1 day before injection. Representative images at day 10 and day 24 after injection show the presence of signal-emitting cells in the thorax (right); temporal increases in bioluminescence (*14*) were quantified in relative luminescence units (left; *n* = 5 mice; error bars represent SD). (**F**) Axial (top) and coronal (bottom) MRI images of a mouse from (E) maintained on doxycycline for 12 weeks show solid nodules in the lung. A corresponding bioluminescence image is shown on the right.
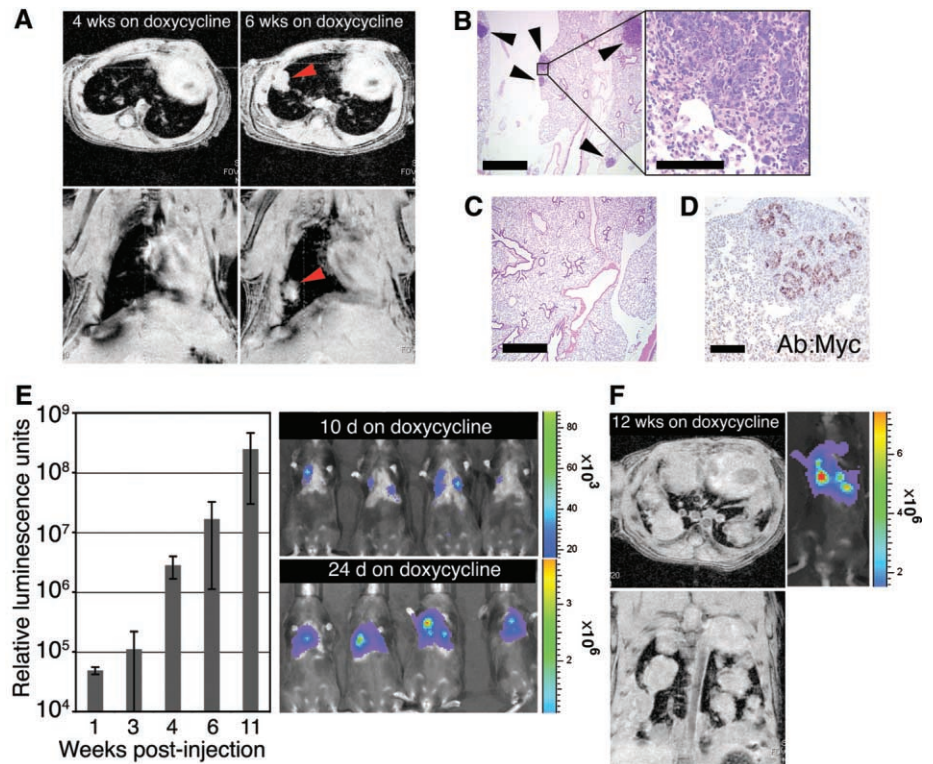
Fig. 2. Delay in oncogene activation does not preclude the development of ectopic mammary tumors. (**A**) Bioluminescence in *Rag1*[−/−] mice after IV delivery of 1×10^5 mammary cells from doxycycline-naïve *TOMT:IRES:Luc;MTB* mice is undetectable before doxycycline exposure but can be induced at various times after placing mice on doxycycline 1.5 (◆, *n* = 7 mice), 8 (■, *n* = 3 mice), or 17 weeks (●, *n* = 2 mice) after IV injection. Downward arrows indicate times of addition of doxycycline to the diet. Error bars represent SD. Representative bioluminescence images (right) obtained 10 days after injection in the absence of doxycycline (top right) and after 2 additional weeks on doxycycline (bottom right). (**B**) Histologically similar metastatic tumors in lungs of *Rag1*[−/−] mice after IV delivery of 1×10^5 mammary cells from doxycycline-naïve *TOMT:IRES:Luc;MTB* mice exposed to doxycycline for 8 weeks starting 10 days after (left) or 1 day before (right) IV injection. Scale bars, 1 mm (top) and 0.1 mm (bottom). (**C**) Mitotic activity in tumor foci in lung sections from *Rag1*[−/−] recipient of *TOM;TOR;MTB* cells (left; stained with anti-pH3) or from *Rag1*[−/−] recipient of *TOMT:IRES:Luc;MTB* cells (right; stained with anti-BrdU serum after BrdU labeling) (*14*). Scale bar, 0.1 mm. (**D**) Angiogenic proficiency demonstrated by perfusion of the ectopic tumor with biotinylated lectin (left) (*14*) and by staining endothelial cells within tumor foci with anti-CD31 serum in a *Rag*[−/−] recipient of the *TOMT:IRES:Luc; MTB* cells (right). Scale bar, 0.1 mm.
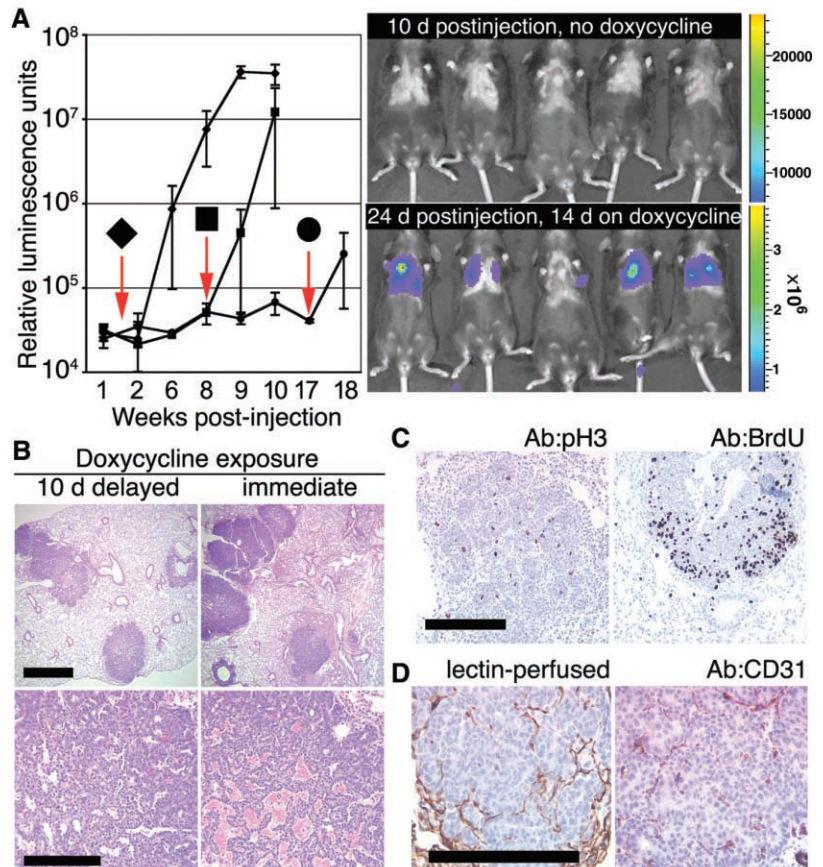
To investigate whether mammary cells from these mice can be induced to form metastasis in the absence of transformation at the primary site, we modified the traditional experimental metastasis assay (14). In the modified approach, instead of IV delivery of tumor cells from doxycycline-treated animals into new recipients, we injected dissociated morphologically normal mammary cells from mature *TOM;TOR;MTB* animals never exposed to doxycycline into the tail veins of *Rag1*⁻/⁻ (15) females on a doxycycline diet. In this way, the injected cells can become transformed only in the bloodstream or tissues of the recipient mouse, an experimental situation that has not been previously examined. Magnetic resonance imaging (MRI) was used to survey *Rag1*⁻/⁻ recipients for evidence of tumor foci in the lungs. Solitary nodules were observed in four out of four recipients 6 weeks after injection (Fig. 1A), and histological sections showed foci of mam-

mary adenocarcinoma in the lung (Fig. 1B); in contrast, no tumors were found in control mice that did not receive doxycycline (Fig. 1C and table S1). The experimental metastases were histologically identical to the spontaneous metastases observed in tumor-bearing *TOM;TOR;MTB* mice (fig. S1) as well as to the primary mammary tumors arising in donor animals (13), and pulmonary nodules contained cells positive for MYC (Fig. 1D), keratin 8 (K8), smooth-muscle actin (SMA), and keratin 6 (K6) (fig. S2). These findings demonstrated that the tumorigenic capacity conferred on mouse mammary cells by coexpression of *MYC* and *Kras*^D12^ can be realized in the ectopic environment of the lung.

We confirmed this observation in a different mouse model, *TetO-PyMT:IRES:Luc;MMTV-rtTA* (*TOMT:IRES:Luc;MTB*), recently generated in our lab (14). Expression of the polyoma middle T (*PyMT*) oncogene in this line is mammary gland–

specific and doxycycline-dependent and can be monitored through the coordinate expression of the reporter gene (*Luc*) encoding firefly luciferase (fig. S3). When dissociated mammary cells from mature *TOMT:IRES:Luc;MTB* animals never exposed to doxycycline were injected intravenously in *Rag1*⁻/⁻ females on a doxycycline diet, a bioluminescence signal was apparent over the thorax of recipient mice within 2 weeks (Fig. 1E). The development of mammary tumors in the lungs of injected mice was documented by MRI imaging (Fig. 1F) and by histology (Fig. 2B). Ectopic mammary tumor foci derived from either *TOM;TOR;MTB* or *TOMT:IRES:Luc;MTB* transgenic lines displayed characteristics associated with oncogene activation at the primary site, including robust proliferation (Fig. 2C) and dense vasculature (Fig. 2D). Therefore, upon activation of potent oncogenes, previously untransformed mouse mammary cells delivered to the systemic circulation produce metastatic-like disease in the pulmonary parenchyma without having undergone transformation at the primary site.

To explore whether untransformed mammary cells can survive in the bloodstream and in the ectopic environment of the lung without oncogene expression yet still be induced to develop tumors at a later time, we injected dissociated mammary cells from *TOMT:IRES:Luc;MTB* animals never exposed to doxycycline into the lateral tail veins of *Rag1*⁻/⁻ females on a doxycycline-free diet. No bioluminescence signal was observed in the lungs of recipient mice in up to 4 months of monitoring on this diet (Fig. 2A). When recipients were instead placed on doxycycline 1.5, 8, or 17 weeks after IV injection of transgenic mammary cells, we detected bioluminescence in the chest within 2 weeks of the start of doxycycline exposure (Fig. 2A). The histological appearance and the total number of foci in the lungs were similar in animals placed on a doxycycline diet 1 day before IV injection and 10 days after the injection [32 ± 6 (SD), *n* = 5 mice and 26 ± 6 (SD), *n* = 4 mice, respectively] (Fig. 2B). Conversion to malignancy at later times of induction was not measured because expression of the *MMTV-rtTA* transgene in the *MTB* line becomes non-uniform with age (16). These observations show that cells responsible for development of the ectopic mammary tumors can persist in the lung for up to 17 weeks in the absence of oncogene expression.

To rule out the possibility that low-level or transient expression of the transgenic oncogene(s) occurs in the absence of doxycycline, we carried out the modified experimental metastasis assay with mammary gland preparations from animals lacking any transgenic oncogenes (C57BL6/J mice) or from those expressing a gene encoding the "enhanced" green fluorescent protein (GFP) from the chicken β-*actin* promoter (β-*actin-GFP*) (17). Three weeks after the injection of mammary cells from β-*actin-GFP* mice, green foci were observed with fluorescent microscopy of the whole lungs in all recipients (Fig. 3A). These foci lacked the nodular appearance of metastatic tumors and
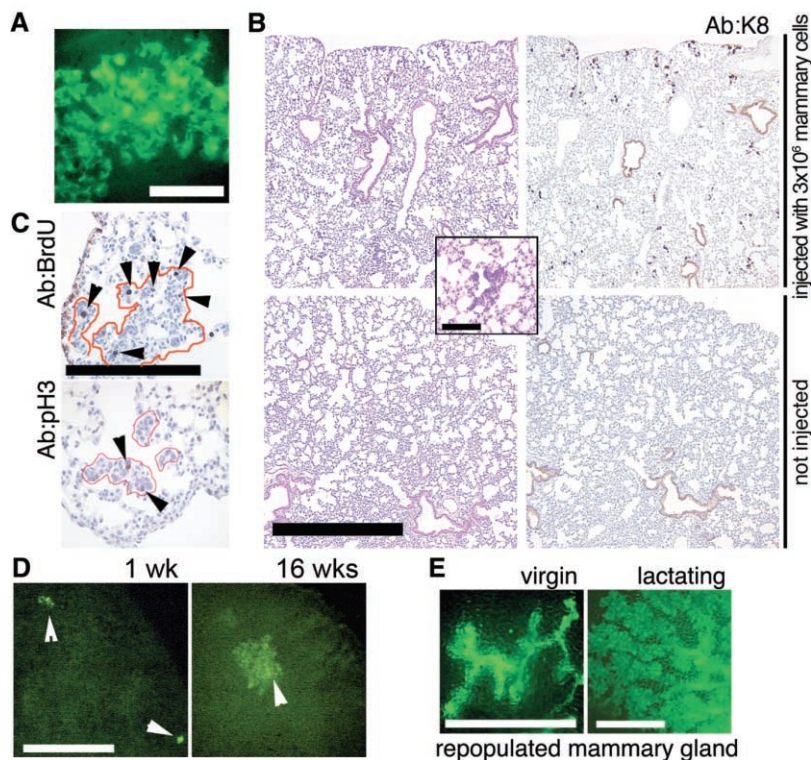
Fig. 3. Mammary cells without an oncogenic transgene can persist in the lung. (**A**) Focus of green cells observed under excitation light (14) in a whole fresh lung of a *Rag1*⁻/⁻ recipient 3 weeks after IV injection of 5 × 10⁵ dissociated mammary cells from a β-*actin-GFP* mouse. Scale bar, 0.1 mm. (**B**) Representative size and distribution of the H/E–stained ectopic foci in lung sections from *Rag1*⁻/⁻ mice injected intravenously with 3 × 10⁶ mammary cells from a β-*actin-GFP* donor at 3 weeks after injection (top left). Inset shows a representative H/E–stained focus at high magnification. A consecutive section (top right) was stained with rat antiserum against K8. No foci were detected by H/E (bottom left) or K8 staining (bottom right) of lung sections from uninjected *Rag1*⁻/⁻ mice. Scale bar, 1 mm. Inset scale bar, 0.2 mm. (**C**) Mitotic activity in ectopic epithelial outgrowths (outlined in red) demonstrated by BrdU-labeling detected with rat anti-BrdU serum (arrowheads, top) or by staining with anti-pH3 (arrowheads, left). Scale bar, 0.2 mm. (**D**) Larger foci of green fluorescent cells were observed by whole-lung imaging under excitation light (14) at 16 weeks after injection (right) as compared with 1 week after injection (left). *Rag1*⁻/⁻ recipients were injected with 3 × 10⁵ dissociated mammary cells from a β-*actin-GFP* mouse. Scale bar, 1 mm. (**E**) Mammary gland repopulation in secondary *Rag1*⁻/⁻ recipients produces a green fluorescent mammary tree detectable under excitation light in the whole-mount preparations 4 or 7 weeks after transplantation (14). Glands were harvested from virgin recipients (left), or host animals were mated and the transplanted glands harvested 1 day postpartum (right). Scale bar, 1 mm.

were inconspicuous on routine histological inspection (Fig. 3B, top left). Staining with antibody to K8 facilitated the detection of the ectopic foci and confirmed their epithelial origin, whereas in the lungs of uninjected mice only the bronchial epithelium was stained (Fig. 3B, right).

To determine whether the ectopic foci of normal epithelial cells persist and grow in the foreign environment of the lung, we counted the total number of discrete foci in lung sections at different times after injection and looked at proliferation markers in these foci. The total number of foci found in lung sections from C57BL6/J recipients injected with 4×10^5 syngeneic mammary cells was similar in the animals surveyed at 3 weeks ($n = 3$ mice) and those surveyed at 10 weeks ($n = 3$ mice) after injection (42 ± 7 and 56 ± 22 in 10 paraffin lung sections, respectively). Moreover, the efficiency with which the wild-type cells were able to form these small epithelial clusters was similar to the efficiency with which we were able to induce ectopic tumors after injecting cells from doxycycline-naïve *TOM;TOR;MTB* donors [1.2 ± 0.4 (SD) versus 1.7 ± 1.4 (SD) per 10,000 cells injected, $n = 6$ and 8 mice, respectively; measured as described in (*14*)]. This result strongly argues that most or all of the mammary cells that are capable of surviving in the lung are able to respond to the initiating oncogene expression by forming an ectopic mammary tumor.

In both nontransgenic C57BL6/J– and β-*actin-GFP*–derived foci, occasional cells displayed mitotic activity (Fig. 3C). Consistent with this result, the green foci found under excitation light in the lungs of animals injected with mammary cells from β-*actin-GFP* mice 16 weeks after injection were larger in size than those found in recipients of the same preparation 1 week after injection (Fig. 3D). Ectopic epithelial outgrowths contained K8- and SMA-positive cells, such as observed in intact mammary glands (fig. S4A), and the outgrowths occasionally displayed a glandular appearance. Despite prolonged residence in the lung (up to 4 months), the green cells recovered from the recipients' lungs were competent to form hollow acinar structures in three-dimensional morphogenesis assays (fig. S4B) and secondary mammary outgrowths in cleared fat pads of *Rag1*$^{-/-}$ females (Fig. 3E). These findings establish that the ectopic cells residing in the lungs are indeed of mammary origin, that they are viable and mitotically active, and that at least some of them are multipotent and able to support full mammary development.

The experiments described here show that, in the absence of an active oncogene, dissociated cells from an untransformed mouse mammary gland can establish residence in the ectopic environment of the lung, grow slowly, and remain clinically undetectable after IV injection. The same cells can give rise to metastatic malignancies upon activation of oncogenes that can produce mammary tumors in an intact gland. It is widely acknowledged that multiple steps are required to establish metastases, including intravasion of cells from primary tumors into blood vessels or lymphatics; survival in the circulation, extravasation, and establishment of cells at ectopic sites; and malignant growth. Because we have injected mammary cells from transgenic mouse donors into tail veins of recipient mice, we have not examined the requirements for intravasation. We have, however, demonstrated that activated oncogenes and cellular transformation are not required for any of the subsequent steps, save for malignant growth at ectopic sites. These findings indicate that properties inherent in normal cells are sufficient for negotiating a substantial portion of the metastatic cascade. Considerable experimental and clinical evidence favors the idea that cells from small cancers may spread to distant sites early in tumorigenesis and account for dormancy and late relapse in human breast cancer (*2*, *18*). Although we do not know whether premalignant cells can enter the systemic circulation during these early stages and become sources of later metastatic tumors, our observations argue that this hypothesis should be tested. The finding that metastatic disease can arise from untransformed mammary cells in the circulation refines our conception of cancer progression, and suggests that each step in the metastatic cascade should be examined to establish its functional requirements, including those performed by normal cells. Such functions might be susceptible to inhibitory strategies that can ablate disseminated pre-malignant or malignant cells and thereby diminish the mortality caused by cancer.

References and Notes
1. D. Hanahan, R. A. Weinberg, *Cell* **100**, 57 (2000).
2. Y. Husemann *et al.*, *Cancer Cell* **13**, 58 (2008).
3. G. P. Gupta *et al.*, *Nature* **446**, 765 (2007).
4. P. B. Gupta *et al.*, *Nat. Genet.* **37**, 1047 (2005).
5. K. A. Hartwell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18969 (2006).
6. M. Jechlinger *et al.*, *J. Clin. Invest.* **116**, 1561 (2006).
7. R. S. Muraoka *et al.*, *J. Clin. Invest.* **109**, 1551 (2002).
8. Y. Kang *et al.*, *Cancer Cell* **3**, 537 (2003).
9. A. Muller *et al.*, *Nature* **410**, 50 (2001).
10. J. Yang *et al.*, *Cell* **117**, 927 (2004).
11. A. E. Karnoub *et al.*, *Nature* **449**, 557 (2007).
12. R. Bernards, R. A. Weinberg, *Nature* **418**, 823 (2002).
13. K. Podsypanina, K. Politi, L. J. Beverly, H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5242 (2008).
14. Materials and methods are available as supporting material on *Science* Online.
15. P. Mombaerts *et al.*, *Cell* **68**, 869 (1992).
16. E. J. Gunther *et al.*, *FASEB J.* **16**, 283 (2002).
17. M. Okabe, M. Ikawa, K. Kominami, T. Nakanishi, Y. Nishimune, *FEBS Lett.* **407**, 313 (1997).
18. J. A. Aguirre-Ghiso, *Nat. Rev. Cancer* **7**, 834 (2007).
19. We thank M. A. Melnick, G. Sanchez, A. Giannakou, and J. Demers for expert handling of the mouse colony; L. Chodosh for providing *MMTV-rtTA* transgenic mice; D. Felsher and J. M. Bishop for providing *TetO-MYC* transgenic mice; A. Olshen for assistance with statistical analysis; and L. K. Tan for assistance with histological analysis. Supported in part by awards from NIH (K01 CA118731 to K.P., P01 CA94060 to H.V., and R24 CA83084 and P30-CA 08748, which provides partial support for core facilities used in conducting this investigation), the Martell Foundation (to H.V.), and the U.S. Department of Defense (W81XWH-05-1-0220 to M.J.).

The Coevolution of Cultural Groups and Ingroup Favoritism

Charles Efferson,[1,2]* Rafael Lalive,[3] Ernst Fehr[1,4]

Cultural boundaries have often been the basis for discrimination, nationalism, religious wars, and genocide. Little is known, however, about how cultural groups form or the evolutionary forces behind group affiliation and ingroup favoritism. Hence, we examine these forces experimentally and show that arbitrary symbolic markers, though initially meaningless, evolve to play a key role in cultural group formation and ingroup favoritism because they enable a population of heterogeneous individuals to solve important coordination problems. This process requires that individuals differ in some critical but unobservable way and that their markers be freely and flexibly chosen. If these conditions are met, markers become accurate predictors of behavior. The resulting social environment includes strong incentives to bias interactions toward others with the same marker, and subjects accordingly show strong ingroup favoritism. When markers do not acquire meaning as accurate predictors of behavior, players show a markedly reduced taste for ingroup favoritism. Our results support the prominent evolutionary hypothesis that cultural processes can reshape the selective pressures facing individuals and so favor the evolution of behavioral traits not previously advantaged.

A cultural group is a group of people who share a set of beliefs, behavioral norms, and behavioral expectations that is recognizably different from those of other groups (*1*). Beliefs, norms, and expectations, however, are often not directly observable, and so by themselves they do not provide a practical basis for identifying cultural groups in everyday social interactions. Nonetheless, cultural groups are frequently identifiable through ethnic markers, which are arbitrary but observable traits like dress, dialect, and body modification that symbolically and conspicuously signal group affiliation (*1*–*5*).

Symbolic traits of this sort can be crucial to social and economic outcomes. When ethnic markers covary with other cultural traits, individuals can

potentially use markers to everyone's mutual advantage as indicators of what would otherwise be unobservable variation in beliefs, norms, and expectations. More nefariously, ethnic markers can lead to segregation, ethnic discrimination, and persistent inequality, even in the paradoxical cases when everyone prefers integration (6–8) or when ethnicity indicates nothing about competence in a given domain (9, 10). Indeed, parochialism and prejudice often mar intergroup relations. People show favoritism toward ingroup members and indifference, hostility, or mistrust toward outgroup members (11–19). They do so even when groups are transient and group boundaries rest on the flimsiest of distinctions among individuals (15, 20–22). These findings have potentially broad significance because recent theoretical research has closely and surprisingly tied outgroup hostility to the evolution of human prosociality within groups (23, 24).

None of this, however, explains how a group gets to be a group and why. The long tradition of empirical research on intergroup relations (11–22, 25) includes two basic approaches to defining groups. Studies have either used preexisting cultural groups, which formed beyond the ken of the studies in question, or subjects were assigned to groups exogenously as part of an experiment involving the effects of social categorization. These methods can be powerful for many questions (12, 16), but they cannot expose the mechanisms behind the formation of cultural groups. These mechanisms also represent a gap in evolutionary theories of human prosociality. Although the initial evolution of cultural groups may have little to do with cooperation, much of the theory on the evolution of human prosociality relies heavily on the observation that human populations are subdivided into cultural groups (23, 24, 26). This theoretical work, however, simply imposes the required population structure exogenously. The endogenous formation of cultural groups represents a plausible route to the required population structure that figures prominently but remains unexplained in evolutionary theories of human prosociality.

We conducted a set of experiments to identify the conditions required for cultural groups to form endogenously and for subjects to show ingroup favoritism in their subsequent social interactions. We used neither preexisting cultural groups nor groups created exogenously by the experimenter. Our task instead was to see if and when symbolically marked groups form endogenously and whether their formation can lead to a preference for interactions with others having the same symbolic marker. This preference was our operational measure of ingroup favoritism in the experiment, and more generally such preferences

[1]Institute for Empirical Research in Economics, University of Zürich, Blümlisalpstrasse 10, 8006 Zürich, Switzerland. [2]Santa Fe Institute, NM 87501, USA. [3]Department of Economics, University of Lausanne, 1015 Lausanne, Switzerland. [4]Collegium Helveticum, 8092 Zürich, Switzerland.

*To whom correspondence should be addressed. E-mail: efferson@iew.uzh.ch

can limit social interactions across cultural boundaries and potentially play a key role in the development of ethnocentric attitudes (27). If such a preference were to emerge endogenously in our setting, the result would support a central hypothesis in evolutionary social science (27–31). This hypothesis posits that a cultural evolutionary process can modify the selective environment facing individuals and so lead to the evolution, whether cultural or genetic, of traits that were not previously advantageous. In our case, the question is whether the evolution of cultural groups during an experiment can reconstitute the social environment to benefit ingroup favoritism in a way that did not obtain at the beginning of the experiment.

In theory, cultural groups form when variation in an unobservable but socially critical variable becomes manifest. Consider a population of players playing a simultaneous, two-person coordination game with multiple equilibria. Players can choose behavior A or B. If two players meet and choose the same behavior, a large payoff results. If they choose different behaviors (32), a small payoff results. Some players expect to coordinate on A, others on B. If players with different expectations meet, an information problem results. One simply has to play the odds and risk miscoordinating with someone who has incompatible expectations. This kind of problem is general. Variation in behavioral norms and expectations is widespread (1, 33, 34), and the mixing of people with different expectations occurs frequently (1, 35, 36). This mixing, however, creates the potential for people with discordant social expectations to meet, interact, and miscoordinate. Variation in expectations, however, is not enough for the existence of cultural groups because this variation is not directly observable.

Symbolic markers can change matters greatly, but only if they covary with expectations and by extension behavior. To illustrate, let players in our coordination game wear shirts with either
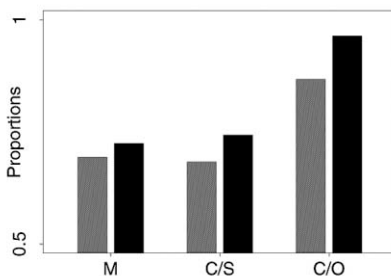
Fig. 1. Summary of linked choices for the marker-randomized (gray) and marker-maintained (black) treatments. The behavior and marker chosen in stage 1 are coded as either linked or unlinked relative to the behavior and marker chosen in stage 1 of the previous period. Proportions are plotted for the cases in which the player miscoordinated (M) in the previous period, coordinated on the suboptimal (C/S) behavior (i.e., A in subpopulation 2 or B in subpopulation 1), and coordinated on the optimal (C/O) behavior (i.e., A in subpopulation 1 or B in subpopulation 2).

triangles or circles. The shape on one's shirt does not affect payoffs, and so it fills the theoretical role of a symbolic marker. Consider a hypothetical population of 100 people, 50 of whom expect to coordinate on A and 50 on B. In addition, the 50 players who expect to coordinate on A have triangles on their shirts, and the 50 players who expect to coordinate on B have circles. The distribution of behavior-marker types in the population is consequently 50 (A, ▲) individuals, 0 (A, ●) individuals, 0 (B, ▲) individuals, and 50 (B, ●) individuals. The covariation between behavior and marker is at its maximum possible value in this example, and the markers perfectly reveal expectations and their associated behaviors in the coordination game. More generally, when covariation characterizes the distribution of behavior-marker types, the observable markers allow one to draw statistical inferences about what is unobservable but really important, namely, behavioral expectations in a social setting with multiple equilibria. When this is true, interacting preferentially with others having the same marker reduces the probability of miscoordination and
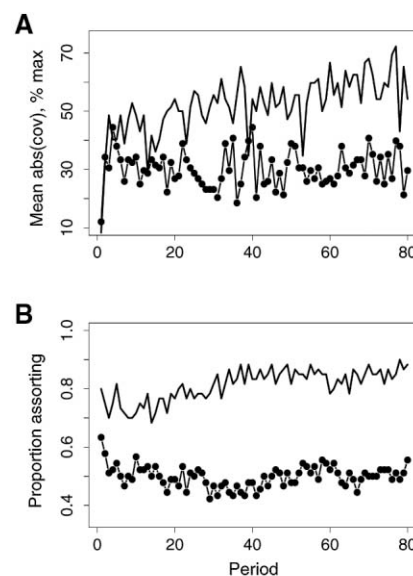
Fig. 2. (**A**) The informational content of the marker. The graph shows the mean magnitude of the covariance between behavior and marker in a subpopulation relative to the theoretical maximum for the marker-randomized (line with filled circles) and the marker-maintained (solid line) treatments. The period trend for marker-randomized is not significant [Newey-West (40) regression, maximal lag of 10, t test, P = 0.368], whereas it is highly significant for the marker-maintained treatment (Newey-West, lag of 10, t test, P < 0.001). (**B**) Ingroup favoritism, as indicated by the proportion of players requesting a partner with the same shape. The marker-randomized period trend is not significant (Newey-West, lag of 10, t test, P = 0.868). The marker-maintained period trend is highly significant (Newey-West, lag of 10, t test, P < 0.001), leading to large differences in ingroup favoritism across treatments.

increases expected payoffs. The puzzle, however, is how to get strong covariation endogenously in decentralized societies under limited information about the distribution of behavior-marker combinations. How does symbolic meaning emerge in the absence of fiat? Interestingly, mixing players with different expectations, which creates the original problem, also creates a potential solution. It does so by producing small amounts of covariation (37) that can feed back into the system and accumulate dynamically (38, 39).

The accumulation of covariation requires more than mixing, however, because mixing by itself often creates only a small amount of covariation between behavior and marker (38). During our experiment, individuals did not have information about the aggregate distribution of behavior-marker combinations, and thus it would have been difficult or impossible to recognize an initially weak relation between behavior and marker. Covariation can increase, however, if individuals link behaviors and markers in specific ways. Linkage refers to a tendency for an individual either to retain both her current behavior and marker or to change both her behavior and marker; what an individual does not do is change one trait but not the other. Linkage is crucial because it preserves the covariation created by earlier mixing, while continued mixing creates additional covariation that feeds back into the system and gets added to existing covariation. The result is that the total covariation accumulates, and this increases the economic incentives to interact with others having the same marker. For covariation to accumulate, however, linkage should not be indiscriminate. Rather, theory suggests it should be more prevalent in specific situations like those in which individuals acquire information about economically successful behavior-marker combinations (38, 39). If individuals, however, never link under any circumstances because they choose behaviors and markers independently, covariation is constantly destroyed, and markers cannot become strongly associated with behavior.

We conducted the following experiment to see if players would show a preference for (i) linking behaviors and markers and for (ii) interacting with partners displaying the same marker. In addition, we wanted to know (iii) whether linkage, if present, would generate sizable covariation between behavior and marker, which would then enable subjects to increase coordination via ingroup favoritism. Players were assigned to one of multiple populations of 10. We randomly subdivided these 10 players into two subpopulations of 5. Players within a subpopulation played one of two coordination games (table S1). Each game had two pure-strategy equilibria, and thus players had to solve a co-ordination problem. Both games had two behaviors to choose from, A and B, but in subpopulation 1, coordinating on A (41 points for each player paired with another playing A) was better than coordinating on B (21 points for each player paired with another playing B), whereas in sub-population 2, coordinating on B (41 points) was better than coordinating on A (21 points). Mis-

coordinating in either subpopulation brought a small payoff (1 point). Payoffs were designed to draw players in different subpopulations toward different behaviors and so mimic the variation in norms, preferences, and expectations that often exists because of historical separation or important but unobservable environmental differences.

To create a persistent coordination problem, players from the different subpopulations were mixed, and they were never told to which sub-population they were assigned. If players had remained in their initial subpopulations, the game would have posed little problem. Players would have soon figured out their respective situations, and presumably players in subpopulation 1 would have only chosen A, whereas players in subpopulation 2 would have only chosen B. Each period, however, a randomly selected player from subpopulation 1 and a randomly selected player from sub-

population 2 switched subpopulations. All players knew this would happen, but no one knew which two players had switched. In sum, each player had a strong incentive to develop accurate expectations about her current subpopulation, but from time to time she found herself in a new situation where her social expectations ran askew of local norms.

Players could also condition social interactions on symbolic markers. In each period, each player chose one of two shapes, ▲ or ●. A player's payoff did not directly depend on her shape, but players could use shapes to influence with whom they would play the coordination game (39). The experiment lasted 80 periods. Each period proceeded as follows.

Stage 1. Each player chose a payoff-relevant behavior, A or B, for the coordination game and a payoff-irrelevant shape, ▲ or ●.

Stage 2. An unidentified player from each subpopulation switched subpopulations.
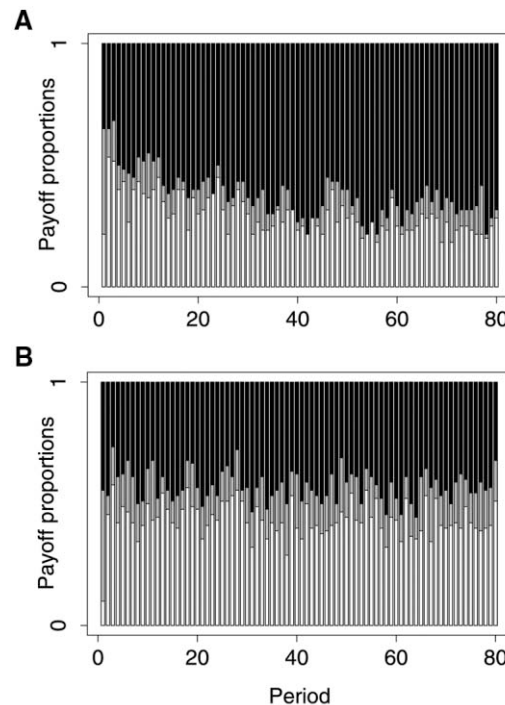
Fig. 3. Payoff proportions in the marker-maintained treatment (**A**) and the marker-randomized treatment (**B**). The graphs show the distribution of players by period coordinating on the optimal behavior (black) given the subpopulation (A in 1, B in 2), coordinating on the suboptimal behavior (gray) given the subpopulation (A in 2, B in 1), or miscoordinating (white). See supporting online text for a multinomial regression analysis.

Fig. 4. Ingroup favoritism for the modified marker-maintained (solid line), payoff-equivalent (dashed line), and fixed-marker (line with open circles) treatments. Newey-West (40) regressions indicate that the modified marker-maintained treatment began with more assortment than the other two treatments, and the differences across treatments increased through time. Comparison of regression results for the modified marker-maintained and payoff-equivalent treatments shows that the intercepts are significantly different (z test, $P < 0.001$), as are the period trends (z test, $P < 0.001$). Comparison of results for modified marker-maintained and fixed-marker shows that both the intercepts (z test, $P < 0.001$) and period trends (z test, $P = 0.019$) are significantly different.
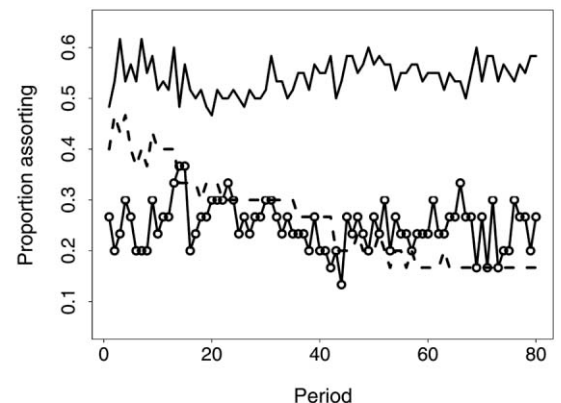
Stage 3. Each player indicated whether she wanted to play the coordination game with (i) a randomly selected player with the same shape from her subpopulation or (ii) any randomly selected player from her subpopulation.

Stage 4. Each player was paired using her choice in stage 3 and received a payoff based on her behavior, her partner's behavior, and their subpopulation.

To clarify our discussion of the results, when there was little or no covariation between behavior and marker, we will call a set of individuals who shared the same marker a "trivial" group. These groups were trivial in the sense that the markers partitioned the population into circles and triangles, but these markers did not reliably reflect any underlying variables affecting payoffs. We will call a group "cultural," in contrast, only when a set of individuals shared the same marker after a sustained increase in the aggregate covariation between behavior and marker. Groups were cultural in this case because markers did not simply partition the population into circles and triangles; they also, on average, partitioned the population into those who expected to coordinate on A versus those who expected to coordinate on B.

The experiment consisted of two treatments. In the marker-randomized treatment, each player was randomly assigned a shape after stage 2 regardless of the shape chosen in stage 1. In the marker-maintained treatment, each player retained her chosen shape. The marker-randomized treatment was a control treatment in which marker randomization precluded the possibility of the marker becoming an accurate predictor of behavior. The comparison between the two treatments shows (i) how much informational content the marker acquired in the marker-maintained treatment beyond the baseline when markers were randomly assigned and (ii) whether any differences in informational content translated into differences in the preference for ingroup favoritism.

Subpopulations were not equivalent to symbolically marked groups, whether trivial or cultural. In a given period, a player's subpopulation was the pool of players available for social interaction. A symbolically marked group, in contrast, was the set of players from the entire population with the same marker. In short, the division of players into two subpopulations, one favoring behavior A and the other behavior B, sustained variation in norms and expectations. This variation, however, was not observable, and so it could not by itself serve as a means of distinguishing one group from another. Symbolic markers, in contrast, were observable traits, and they could serve as a means of distinguishing one group from another. Markers, however, did not bear any necessary relation to behavior and subpopulation. The significance of markers, in essence, could only emerge during the experiment as a result of player choices. Markers had the potential to become the basis for determining cultural group affiliation ex post, and indeed that was our question, but they were devoid of content ex ante.

For a sustained increase in covariation, individuals have to link the behavioral and marker

dimensions. We coded behavior-marker choices from stage 1 of periods 2 to 80 as "linked" or "unlinked." A linked choice was one in which a player either retained her behavior and chosen marker from the previous period or changed both. An unlinked choice was when she changed her behavior or marker but not both. A strong preference toward linked choices was present in general (Fig. 1), but it was significantly stronger in the specific case when a player received the optimal coordination payoff in the previous period (conditional logit, $P < 0.001$, table S3). These linked choices consisted almost exclusively of choices in which the player retained her behavior and marker from the previous period (figs. S1 and S2). In addition, although the preference for linked choices after coordinating on the optimal behavior was present in both treatments, it was significantly stronger in the marker-maintained treatment (conditional logit, $P < 0.001$, table S3). These results indicate that players showed a general tendency to couple behaviors and markers. This tendency, however, was strongest when a player hit upon a successful behavior-marker combination, and it was further reinforced and amplified in the marker-maintained treatment when the marker was not prevented from acquiring meaning.

Substantial linkage at the individual level produces covariation between behavior and marker at the aggregate level. If strong enough and specific enough, the linkage exhibited in the experiment should have produced a significant increase in covariation in the marker-maintained treatment, when it was possible, but not in the marker-randomized treatment, when it was not. Even though linkage was present, however, covariation should have been similar in the two treatments at the beginning of the experiment, before covariation had time to accumulate. Only in later periods should the covariation have been significantly higher in the marker-maintained treatment. The aggregate covariation between behavior and marker indeed followed this dynamical pattern. During the first five periods, the covariation was not different in the two treatments (Welch two-sample t test, df = 7.01, two-sided P = 0.68, Fig. 2A), whereas in the final five periods, the covariation was significantly higher in the marker-maintained case (Welch two-sample t test, df = 12.107, two-sided $P < 0.001$, Fig. 2A). Covariation thus strongly and significantly increased in the marker-maintained treatment but not in the marker-randomized case. This led to a strong overall treatment difference in the accumulation of the markers' predictive power [Newey-West regression (40), $P < 0.001$, Fig. 2A].

The presence of covariation does not mean that players will exploit it by assorting into groups characterized by shared markers. Players could simply fail to recognize the association between behavior and marker as it developed, or they could fail to recognize its usefulness. Nonetheless, players exhibited an increasing inclination to request partners with the same shape as covariation accumulated. Throughout the marker-randomized treat-

ment, players requested same-shape partners roughly 50% of the time (Fig. 2B), a result consistent with indifference concerning the two interaction policies. In the marker-maintained treatment, however, players increasingly requested partners having the same shape as time passed. This increase was highly significant (Newey-West regression, $P < 0.001$), and the vast majority of players (87%) requested partners with the same shape in the final five periods (Fig. 2B), indicating that ingroup favoritism became an almost universal phenomenon.

In the presence of covariation, this kind of ingroup favoritism should lead to more coordination and improved payoffs, but the strength of the effect will vary with the degree of covariation and preferential assortment. A calculation of the mean payoff over periods for each subject shows that payoffs were significantly different across the two treatments. The mean payoff in the marker-randomized treatment was 20.819 points, and it was 27.454 in the marker-maintained treatment (Welch two-sample t-test, df = 88.912, two-sided $P < 0.001$). This difference, however, depended specifically on the dynamical increase in the markers' predictive content in the marker-maintained treatment, and this fact is central to our finding that the evolution of cultural groups changed the incentives associated with ingroup favoritism. Specifically, for those players who requested a partner with the same shape, the mean payoff per period was not significantly different between the two treatments in the first five periods (Welch two-sample t-test, df = 123.139, two-sided $P = 0.1638$), whereas it was highly significant in the final five periods (Welch two-sample t test, df = 105.733, two-sided $P < 0.001$). The higher overall payoffs in the marker-maintained treatment stemmed from an increase in coordinating on the optimal behavior in each of the two subpopulations (Fig. 3). A detailed analysis formally confirms the substantial and robust payoff effect that resulted from assorting on markers in the marker-maintained treatment (37).

These results show how the evolution of cultural groups can reconstitute the social environment and produce selection for an ingroup bias that was not initially advantageous. If selective pressures of this sort were common in past human societies, a plausible outcome would arguably be a relatively inflexible bias leading individuals to prefer others similar in some symbolic dimension. This idea is consistent with much research showing an astonishing willingness for subjects to exhibit ingroup favoritism when groups are based on trivial, short-lived distinctions (12, 15, 16, 21, 22). For our study, this could mean that the marker-based assortment we documented largely reflected a readiness to favor the ingroup that was already in place when the subjects came to the lab, and it did not stem from the endogenous formation of cultural groups during the experiment. In particular, although we found a pronounced difference in assortment dynamics in our two treatments, we still found a strong tendency to assort in the marker-randomized treatment. This assortment was relatively meaningless with respect to payoffs, but

because requesting a partner with the same shape was free, it is consistent with two different motives on the part of players. Players could have simply been indifferent between two largely meaningless, cost-free social interaction policies, or they could have had a strong residual taste for same-shape partners even when such pairings did not improve payoffs. To distinguish between these two possibilities, we conducted a second experiment with three treatments, all of which required subjects to pay a small cost for ingroup favoritism.

In the three treatments of our second experiment (37), subjects had to pay a cost of 1 point when they requested and were successfully paired with a partner having the same shape. To maximize the salience of the marker, all players retained their chosen markers in all treatments. In the payoff-equivalent treatment, the payoff structure was changed such that coordinating on A or B yielded the same payoff (21 points) regardless of the players' subpopulation. Because the payoff structure in the subpopulations was identical in this case, players did not differ in terms of some unobservable variable related to payoffs, and thus they had no material problem the markers could help them solve. They could, of course, continue to bias their interactions toward those having the same marker if willing to pay the cost. In the fixed-marker treatment, players only chose a marker in the first period. This marker was then retained for all 80 periods. In this treatment, markers were ostensibly similar to traits like race that are often perceived as immutable. Because of this perceived immutability, which may or may not be an accurate perception, such traits are especially prone to essentialist generalizations and are thus prime candidates for generating ingroup favoritism and outgroup hostility (41). A truly immutable marker, however, like the one we implemented, should not evolve to be a stable predictor of behavior because individuals cannot adjust their markers to reflect changing social circumstances. In the fixed-marker treatment, for example, players could benefit from changing their expectations about where to coordinate when they changed subpopulations, but they could not change their markers to signal their shifting expectations. Players could nonetheless choose to assort on marker, if they wished. Lastly, as a new baseline, the modified marker-maintained treatment was similar to the original marker-maintained treatment, but it involved the same assortment cost used in the payoff-equivalent and fixed-marker cases.

As in the original marker-maintained treatment, the covariance between behavior and marker accumulated at a significant rate through time in the modified marker-maintained treatment [Newey-West (40) regression, maximal lag of 10, period trend t test, $P = 0.003$]. In early periods, the covariance in the fixed-marker treatment was lower than it was in the modified marker-maintained treatment, and this difference was marginally significant (z test on Newey-West estimated intercepts, $P = 0.067$). Furthermore, unlike the modified marker-maintained case, covariation did not accu-

mulate through time in the fixed-marker treatment (Newey-West regression, lag of 10, period trend t-test, $P = 0.294$). The estimated time trend was slightly negative, and this was significantly different from the positive trend in the modified marker-maintained treatment (z test on Newey-West estimated period coefficients, $P = 0.002$). In the payoff-equivalent treatment, covariance was significantly lower in early periods than it was in the modified marker-maintained treatment (z-test on Newey-West estimated intercepts, $P < 0.001$), and it declined even further at a significant rate (Newey-West regression, lag of 10, period trend t test, $P < 0.001$). In this case, covariance actually declined all the way to 0 because all players soon converged on A in all subpopulations. With no variation in behavior, covariation between behavior and marker is not possible. Shared history was sufficient to form accurate expectations about where to coordinate, and the marker was not useful in this respect. In sum, trivial groups became cultural groups in the modified marker-maintained treatment, but trivial groups remained trivial in the payoff-equivalent and fixed-marker treatments.

Players, in turn, responded strongly to the resulting variation in the accumulated predictive power of markers. In the modified marker-maintained baseline, roughly 55 to 60% of the players requested partners with the same shape in later periods (Fig. 4). In the payoff-equivalent and fixed-marker treatments, however, only 15 to 25% assorted on shape in later periods, and the differences relative to the baseline were highly significant (Fig. 4). The payoff-equivalent case is especially clear because, as mentioned above, all players eventually played A in both subpopulations, and the predictive value of the markers went to zero as a result. Correspondingly, the proportion of players requesting same-shape partners unraveled relentlessly as the experiment progressed (Fig. 4). The fact that assortment did not disappear altogether suggests that perhaps a few players had a weak taste for assortment even when this did not improve coordination. Altogether, however, our results show that the preference for interacting with similarly marked players varied strongly according to whether markers became accurate predictors of behavior in the face of heterogeneous behavioral expectations. In short, ingroup favoritism had little to do with an unconditional preference for similarly marked partners and a lot to do with whether trivial groups evolved into cultural groups. For this cultural evolutionary transition to happen, two requirements had to be met. First, players had to differ persistently in some important but unobservable dimension that could sustain symbolic representation. Our payoff-equivalent treatment removed this feature, and assorting on shape steadily declined through time. Second, the symbolic markers themselves had to be freely chosen and mutable in a way that allowed an association between markers and unobservables to develop. Our marker-randomized and fixed-marker treatments removed this feature, and assorting on shape was relatively low in all periods when

compared to their respective marker-maintained treatments.

The research on intergroup processes indicates that people have a willingness to show ingroup favoritism, and in particular this holds even when groups are trivial and evanescent (12, 13, 16–18, 20–22, 25). This research tradition has generally examined neither the evolutionary mechanisms behind group formation nor the impact of these mechanisms on ingroup favoritism. We implemented an experiment in which the significance of groups had to arise, if at all, endogenously, thus providing an evolutionary foundation for ingroup favoritism. In this setting, trivial groups remained trivial under certain circumstances, but under other circumstances they developed into cultural groups composed of individuals who shared both behavioral expectations and symbolic markers signaling group affiliation. Ingroup favoritism was strongly associated with cultural groups but not with trivial groups. Our experiments made exclusive use of coordination games, which serve as a kind of generic proxy for strategic settings with multiple equilibria. Many strategic settings are characterized by multiple equilibria (42), and thus the dynamical processes examined here have potentially broad significance. The mechanisms implicated in the evolution of human prosociality, for example, often produce multiple equilibria (43, 44), and so cooperation is a behavioral domain with considerable scope for the path-dependent evolution of groups with different norms and expectations. In this sense, cooperation can be analogous to coordination. Even more generally, whenever people have a shared interest in distinguishing among themselves in terms of their unobservable information (38), whatever that means in a given situation, the logic behind the evolution of cultural groups holds.

References and Notes
1. F. Barth, Ed., *Ethnic Groups and Boundaries: The Social Organization of Cultural Difference* (Little, Brown, Boston, 1969).
2. D. E. Blom, *J. Anthropol. Archaeol.* **24**, 1 (2005).
3. J. B. Eicher, Ed., *Dress and Ethnicity: Change Across Space and Time* (Berg, Oxford, 1995).
4. C. Silverman, *J. Am. Folkl.* **101**, 261 (1988).
5. C. Torres-Rouff, *Curr. Anthropol.* **43**, 163 (2002).
6. S. Bowles, *Microeconomics: Behavior, Institutions, and Evolution* (Russell Sage, New York, 2004).
7. T. Schelling, *Micromotives and Macrobehavior* (Norton, New York, 1978).
8. H. P. Young, *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions* (Princeton Univ. Press, Princeton, NJ, 1998).
9. R. L. Axtell, J. M. Epstein, H. P. Young, in *Social Dynamics*, S. N. Durlauf, H. P. Young, Eds. (MIT Press, Cambridge, MA, 2001), pp. 191–211.
10. S. Bowles, G. Loury, R. Sethi, "Is equal opportunity enough? A theory of persistent group inequality," Santa Fe Institute Working Paper; www.santafe.edu/~bowles.
11. H. Bernhard, U. Fischbacher, E. Fehr, *Nature* **442**, 912 (2006).
12. J. F. Dovidio, P. Glick, L. A. Rudman, Eds., *On the Nature of Prejudice: Fifty Years After Allport* (Blackwell, Oxford, 2005).
13. R. H. Fazio, J. R. Jackson, B. C. Dunton, C. J. Williams, *J. Pers. Soc. Psychol.* **69**, 1013 (1995).
14. E. L. Glaeser, D. I. Laibson, J. A. Scheinkman, C. L. Soutter, *Q. J. Econ.* **115**, 811 (2000).
15. L. Goette, D. Huffman, S. Meier, *Am. Econ. Rev.* **96**, 212 (2006).

16. M. Hewstone, M. Rubin, H. Willis, *Annu. Rev. Psychol.* **53**, 575 (2002).
17. K. D. Kinzler, E. Dupoux, E. S. Spelke, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12577 (2007).
18. Z. Kunda, *Social Cognition: Making Sense of People* (MIT Press, Cambridge, MA, 1999).
19. T. Yamagishi, N. Jin, T. Kiyonari, *Adv. Group Process.* **16**, 161 (1999).
20. M. B. Brewer, *Pers. Soc. Psychol. Bull.* **17**, 475 (1991).
21. M. Sherif, O. J. Harvey, B. J. White, W. R. Hood, C. W. Sherif, *Intergroup Conflict and Cooperation: The Robbers Cave Experiment* (Institute of Group Relations, Univ. of Oklahoma, Norman, OK, 1961).
22. H. Tajfel, M. G. Billig, R. P. Bundy, C. Flament, *Eur. J. Soc. Psychol.* **1**, 149 (1971).
23. R. Boyd, H. Gintis, S. Bowles, P. J. Richerson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3531 (2003).
24. J.-K. Choi, S. Bowles, *Science* **318**, 636 (2007).
25. J. Correll, G. R. Urland, T. A. Ito, *J. Exp. Soc. Psychol.* **42**, 120 (2006).
26. R. Boyd, P. J. Richerson, *J. Theor. Biol.* **215**, 287 (2002).
27. R. Boyd, P. J. Richerson, *The Origin and Evolution of Cultures* (Oxford Univ. Press, Oxford, 2005).
28. R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (Univ. of Chicago Press, Chicago, 1985).
29. W. H. Durham, Ed., *Coevolution: Genes, Culture, and Human Diversity* (Stanford Univ. Press, Stanford, CA, 1991).
30. F. J. Olding-Smee, K. N. Laland, M. W. Feldman, *Niche Construction: The Neglected Process in Evolution* (Princeton Univ. Press, Princeton, NJ, 2003).
31. P. J. Richerson, R. Boyd, *Not By Genes Alone: How Culture Transformed the Evolutionary Process* (Univ. of Chicago Press, Chicago, 2005).
32. It is not crucial for our purposes that identical actions lead to high payoffs. Instead, the necessity of coordinated choices for a high payoff is important. Thus, if the choice combination A for player 1 and B for player 2 led to a high payoff, then the players would simply need to coordinate on (A, B). Almost every bargaining problem or economic exchange involves some necessity for coordinating expectations and actions.
33. J. G. Jorgensen, *Western Indians* (Freeman, New York, 1980).
34. C. M. Judd, B. Park, in *On the Nature of Prejudice: Fifty Years After Allport*, J. F. Dovidio, P. Glick, L. A. Rudman, Eds. (Blackwell, Oxford, 2005), pp. 123–138.
35. S. Bowles, *Science* **314**, 1569 (2006).
36. R. F. Heizer, in *Handbook of North American Indians: California*, R. F. Heizer, W. C. Sturtevant, Eds. (Smithsonian Institution, Washington, DC, 1978), pp. 690–693.
37. Materials and methods are available as supporting material on *Science* Online.
38. R. Boyd, P. J. Richerson, *Cult. Anthropol.* **2**, 65 (1987).
39. R. McElreath, R. Boyd, P. J. Richerson, *Curr. Anthropol.* **44**, 122 (2003).
40. W. K. Newey, K. D. West, *Econometrica* **55**, 703 (1987).
41. J. T. Jost, D. L. Hamilton, in *On the Nature of Prejudice: Fifty Years After Allport*, J. F. Dovidio, P. Glick, L. A. Rudman, Eds. (Blackwell, Oxford, 2005), pp. 208–224.
42. D. Fudenberg, E. Maskin, *Econometrica* **54**, 533 (1986).
43. R. Boyd, P. J. Richerson, *Ethol. Sociobiol.* **13**, 171 (1992).
44. K. Panchanathan, R. Boyd, *Nature* **432**, 499 (2004).
45. This research was supported by the Swiss National Science Foundation (105312-114107) and is part of the Research Priority Program "Foundations of Human Social Behavior—Altruism versus Egoism" at the University of Zurich. We thank S. Bowles for valuable comments on an earlier version of this article and R. McElreath for helpful insights during the initial stages of the project.

Understanding Overbidding: Using the Neural Circuitry of Reward to Design Economic Auctions

Mauricio R. Delgado,[1] Andrew Schotter,[2] Erkut Y. Ozbay,[3] Elizabeth A. Phelps[4]*

We take advantage of our knowledge of the neural circuitry of reward to investigate a puzzling economic phenomenon: Why do people overbid in auctions? Using functional magnetic resonance imaging (fMRI), we observed that the social competition inherent in an auction results in a more pronounced blood oxygen level–dependent (BOLD) response to loss in the striatum, with greater overbidding correlated with the magnitude of this response. Leveraging these neuroimaging results, we design a behavioral experiment that demonstrates that framing an experimental auction to emphasize loss increases overbidding. These results highlight a role for the contemplation of loss in understanding the tendency to bid "too high." Current economic theories suggest overbidding may result from either "joy of winning" or risk aversion. By combining neuroeconomic and behavioral economic techniques, we find that another factor, namely loss contemplation in a social context, may mediate overbidding in auctions.

An unresolved question in the emerging field of neuroeconomics is whether data from neuroscience can inform economic theory such that it motivates behavioral economic institutional design (*1–4*). In this report, we address this question by taking advantage of our knowledge of the neural circuitry of reward to investigate a puzzling economic phenomenon. Specifically, why do people overbid in auctions? (*5, 6*).

Auctions are an old and widely used method in allocating goods (*7*). Mention of them dates back to Roman times, when spoils of war were sold on the block. Although there are many different types

of auctions, they all share the feature that bidders must determine a bidding strategy (or bid function) to be used in submitting their bid. A bid function for a buyer in an auction is a mapping from the value that the bidder places on the good for sale to the bid chosen. A set of bidding functions is considered to be an equilibrium (Nash equilibrium) if, given the strategy used by one's opponents, no bidder has any incentive to change his or her bidding strategy. One robust finding in experimental auctions is that bidders tend to bid above their Nash equilibrium risk-neutral bid function (*5*); this behavior has been labeled "overbidding" in the economics literature. In other words, given the value of the good for sale they submit bids that are "too high." Two competing explanations for this phenomenon exist. Many scholars have assumed that risk aversion is responsible for this increase in bids, because bidding above one's risk-neutral Nash equilibrium bid function is exactly what risk aversion prescribes

(*5, 6, 8*). Another explanation stems from the ideas that bidders enjoy a "joy of winning" the social competition inherent in an auction (*5, 6*).

The goal of this study is to provide insight into the neural circuitry of experimental auctions and to use this insight to generate and test a behavioral economic approach to understand overbidding. First, we used functional magnetic resonance imaging (fMRI) to examine the neural correlates of winning and losing an experimental auction, while modulating potentially important variables such as type of social competition (auction versus lottery) and type of incentive (money versus points with no monetary value). On the basis of these brain imaging results and our understanding of the neural circuitry of reward, we generated a hypothesis concerning the mechanisms underlying overbidding in experimental auctions. We then tested this hypothesis in a behavioral economic experiment.

In the fMRI study, 17 participants were instructed that they would each be playing two types of games: a two-person auction and a lottery (52 events for each treatment) (*9*). Before participants were scanned, they briefly met their competitor for the auction and were informed that they would be playing an unknown but fixed strategy. In the auction game, participants were assigned a value (V) at the beginning of each trial. These values were drawn from a finite set with equal probability. Participants were asked to choose a bid (b) (the decision phase) and were then informed if they won or lost the auction (the outcome phase). There were four possible V's assigned for the good sold (6, 8, 10, 12) and four options for b (2, 5, 7, 8). The competitor bid according to the Nash equilibrium strategy (V:b equals 6:2, 8:5, 10:7, 12:8). In the money condition, V and b represented dollars, and the participants were informed they would receive a payoff of V minus b if they won that trial and zero if they lost. They would be paid their total winnings from one randomly selected block out of the four

[1]Department of Psychology, Rutgers University, Newark, NJ 07102, USA. [2]Department of Economics, New York University, and Center for Experimental Social Science, New York, NY 10003, USA. [3]Department of Economics, University of Maryland, College Park, MD 20742, USA. [4]Department of Psychology, New York University, New York, NY 10003, USA.

*To whom correspondence should be addressed. E-mail: liz.phelps@nyu.edu

16. M. Hewstone, M. Rubin, H. Willis, *Annu. Rev. Psychol.* **53**, 575 (2002).
17. K. D. Kinzler, E. Dupoux, E. S. Spelke, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 12577 (2007).
18. Z. Kunda, *Social Cognition: Making Sense of People* (MIT Press, Cambridge, MA, 1999).
19. T. Yamagishi, N. Jin, T. Kiyonari, *Adv. Group Process.* **16**, 161 (1999).
20. M. B. Brewer, *Pers. Soc. Psychol. Bull.* **17**, 475 (1991).
21. M. Sherif, O. J. Harvey, B. J. White, W. R. Hood, C. W. Sherif, *Intergroup Conflict and Cooperation: The Robbers Cave Experiment* (Institute of Group Relations, Univ. of Oklahoma, Norman, OK, 1961).
22. H. Tajfel, M. G. Billig, R. P. Bundy, C. Flament, *Eur. J. Soc. Psychol.* **1**, 149 (1971).
23. R. Boyd, H. Gintis, S. Bowles, P. J. Richerson, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3531 (2003).
24. J.-K. Choi, S. Bowles, *Science* **318**, 636 (2007).
25. J. Correll, G. R. Urland, T. A. Ito, *J. Exp. Soc. Psychol.* **42**, 120 (2006).
26. R. Boyd, P. J. Richerson, *J. Theor. Biol.* **215**, 287 (2002).
27. R. Boyd, P. J. Richerson, *The Origin and Evolution of Cultures* (Oxford Univ. Press, Oxford, 2005).
28. R. Boyd, P. J. Richerson, *Culture and the Evolutionary Process* (Univ. of Chicago Press, Chicago, 1985).
29. W. H. Durham, Ed., *Coevolution: Genes, Culture, and Human Diversity* (Stanford Univ. Press, Stanford, CA, 1991).
30. F. J. Olding-Smee, K. N. Laland, M. W. Feldman, *Niche Construction: The Neglected Process in Evolution* (Princeton Univ. Press, Princeton, NJ, 2003).
31. P. J. Richerson, R. Boyd, *Not By Genes Alone: How Culture Transformed the Evolutionary Process* (Univ. of Chicago Press, Chicago, 2005).
32. It is not crucial for our purposes that identical actions lead to high payoffs. Instead, the necessity of coordinated choices for a high payoff is important. Thus, if the choice combination A for player 1 and B for player 2 led to a high payoff, then the players would simply need to coordinate on (A, B). Almost every bargaining problem or economic exchange involves some necessity for coordinating expectations and actions.
33. J. G. Jorgensen, *Western Indians* (Freeman, New York, 1980).
34. C. M. Judd, B. Park, in *On the Nature of Prejudice: Fifty Years After Allport*, J. F. Dovidio, P. Glick, L. A. Rudman, Eds. (Blackwell, Oxford, 2005), pp. 123–138.
35. S. Bowles, *Science* **314**, 1569 (2006).
36. R. F. Heizer, in *Handbook of North American Indians: California*, R. F. Heizer, W. C. Sturtevant, Eds. (Smithsonian Institution, Washington, DC, 1978), pp. 690–693.
37. Materials and methods are available as supporting material on *Science* Online.
38. R. Boyd, P. J. Richerson, *Cult. Anthropol.* **2**, 65 (1987).
39. R. McElreath, R. Boyd, P. J. Richerson, *Curr. Anthropol.* **44**, 122 (2003).
40. W. K. Newey, K. D. West, *Econometrica* **55**, 703 (1987).
41. J. T. Jost, D. L. Hamilton, in *On the Nature of Prejudice: Fifty Years After Allport*, J. F. Dovidio, P. Glick, L. A. Rudman, Eds. (Blackwell, Oxford, 2005), pp. 208–224.
42. D. Fudenberg, E. Maskin, *Econometrica* **54**, 533 (1986).
43. R. Boyd, P. J. Richerson, *Ethol. Sociobiol.* **13**, 171 (1992).
44. K. Panchanathan, R. Boyd, *Nature* **432**, 499 (2004).
45. This research was supported by the Swiss National Science Foundation (105312-114107) and is part of the Research Priority Program "Foundations of Human Social Behavior—Altruism versus Egoism" at the University of Zurich. We thank S. Bowles for valuable comments on an earlier version of this article and R. McElreath for helpful insights during the initial stages of the project.

Understanding Overbidding: Using the Neural Circuitry of Reward to Design Economic Auctions

Mauricio R. Delgado,[1] Andrew Schotter,[2] Erkut Y. Ozbay,[3] Elizabeth A. Phelps[4]*

We take advantage of our knowledge of the neural circuitry of reward to investigate a puzzling economic phenomenon: Why do people overbid in auctions? Using functional magnetic resonance imaging (fMRI), we observed that the social competition inherent in an auction results in a more pronounced blood oxygen level–dependent (BOLD) response to loss in the striatum, with greater overbidding correlated with the magnitude of this response. Leveraging these neuroimaging results, we design a behavioral experiment that demonstrates that framing an experimental auction to emphasize loss increases overbidding. These results highlight a role for the contemplation of loss in understanding the tendency to bid "too high." Current economic theories suggest overbidding may result from either "joy of winning" or risk aversion. By combining neuroeconomic and behavioral economic techniques, we find that another factor, namely loss contemplation in a social context, may mediate overbidding in auctions.

An unresolved question in the emerging field of neuroeconomics is whether data from neuroscience can inform economic theory such that it motivates behavioral economic institutional design (*1–4*). In this report, we address this question by taking advantage of our knowledge of the neural circuitry of reward to investigate a puzzling economic phenomenon. Specifically, why do people overbid in auctions? (*5, 6*).

Auctions are an old and widely used method in allocating goods (*7*). Mention of them dates back to Roman times, when spoils of war were sold on the block. Although there are many different types of auctions, they all share the feature that bidders must determine a bidding strategy (or bid function) to be used in submitting their bid. A bid function for a buyer in an auction is a mapping from the value that the bidder places on the good for sale to the bid chosen. A set of bidding functions is considered to be an equilibrium (Nash equilibrium) if, given the strategy used by one's opponents, no bidder has any incentive to change his or her bidding strategy. One robust finding in experimental auctions is that bidders tend to bid above their Nash equilibrium risk-neutral bid function (*5*); this behavior has been labeled "overbidding" in the economics literature. In other words, given the value of the good for sale they submit bids that are "too high." Two competing explanations for this phenomenon exist. Many scholars have assumed that risk aversion is responsible for this increase in bids, because bidding above one's risk-neutral Nash equilibrium bid function is exactly what risk aversion prescribes

(*5, 6, 8*). Another explanation stems from the ideas that bidders enjoy a "joy of winning" the social competition inherent in an auction (*5, 6*).

The goal of this study is to provide insight into the neural circuitry of experimental auctions and to use this insight to generate and test a behavioral economic approach to understand overbidding. First, we used functional magnetic resonance imaging (fMRI) to examine the neural correlates of winning and losing an experimental auction, while modulating potentially important variables such as type of social competition (auction versus lottery) and type of incentive (money versus points with no monetary value). On the basis of these brain imaging results and our understanding of the neural circuitry of reward, we generated a hypothesis concerning the mechanisms underlying overbidding in experimental auctions. We then tested this hypothesis in a behavioral economic experiment.

In the fMRI study, 17 participants were instructed that they would each be playing two types of games: a two-person auction and a lottery (52 events for each treatment) (*9*). Before participants were scanned, they briefly met their competitor for the auction and were informed that they would be playing an unknown but fixed strategy. In the auction game, participants were assigned a value (V) at the beginning of each trial. These values were drawn from a finite set with equal probability. Participants were asked to choose a bid (b) (the decision phase) and were then informed if they won or lost the auction (the outcome phase). There were four possible V's assigned for the good sold (6, 8, 10, 12) and four options for b (2, 5, 7, 8). The competitor bid according to the Nash equilibrium strategy (V:b equals 6:2, 8:5, 10:7, 12:8). In the money condition, V and b represented dollars, and the participants were informed they would receive a payoff of V minus b if they won that trial and zero if they lost. They would be paid their total winnings from one randomly selected block out of the four

[1]Department of Psychology, Rutgers University, Newark, NJ 07102, USA. [2]Department of Economics, New York University, and Center for Experimental Social Science, New York, NY 10003, USA. [3]Department of Economics, University of Maryland, College Park, MD 20742, USA. [4]Department of Psychology, New York University, New York, NY 10003, USA.

*To whom correspondence should be addressed. E-mail: liz.phelps@nyu.edu

money blocks presented (each encompassing 13 trials) at the end of the study. In the points condition, V and b represented points. Participants were told that the accumulation of points from a random points block (the sum of V minus b for win trials) would be a measure of how well they did relative to other participants, with final anonymous results disseminated at the conclusion of the study. The auction game used the first-price sealed-bid rule in which the participant did not know the V assigned to the competitor on each trial and the higher bid won. In case of identical bids, ties were broken at random. Because losses yielded zero payoffs, a loss did not signify a monetary or points loss per se, but merely that the participant did not win that particular auction (10).

In the lottery game, subjects played against a computer that used the same fixed Nash equilibrium bid strategy as the auction game confederate. Unlike the auction, participants were not required to submit a bid for the random value assigned to them. Rather, they were assigned both V and b at the beginning of each trial and were simply asked to indicate if they wanted to play the lottery for that trial (decision phase). If their assigned b was greater than the b generated by the computer's Nash equilibrium bid, they won the lottery, if not, they lost (outcome phase). As in the auction game, participants played the lottery for either money or points. Behavioral measures of reaction time and choice were collected throughout the experiment, along with postexperimental Likert-scale ratings (9).

As in previous auction studies (5, 6), participants overbid with respect to the risk-neutral Nash equilibrium [$t(16) = 3.04$, $P < 0.008$]. Overbidding compared with choosing the equilibrium bid was greatest when the incentive was monetary [$t(16) = 3.30$ $P < 0.005$]. Overall, participants' chosen b was greater than the equilibrium b on 65% of the money trials and 57% of the point trials (11).

The goal of the fMRI study was to examine the effects of type of social competition (auction versus lottery) and type of incentive (money versus points) on blood oxygen level–dependent (BOLD) responses to winning or losing. Given this, the focus of the analysis was the outcome phase. Statistical maps contrasting wins and losses across all conditions were generated ($P < 0.001$, cluster threshold of 3 mm³ contiguous voxels). Mean beta weights (19) from each region of interest (ROI) defined by this contrast were extracted and input into two separate analyses of variance (ANOVAs) to examine main effects of social competition and incentive during win or loss outcomes separately. Regions identified during the outcome phase included both left and right striatum, specifically the ventral caudate nucleus, previously implicated in monetary outcome processing and learning from feedback (12–18), along with ROIs in the occipital lobe (table S4). An examination of the BOLD response in these ROIs revealed differential responses for type of social competition or for type

of incentive, only in the right and left striatum (Fig. 1A). Activation in this region has previously been shown to be graded according to the magnitude of monetary gain and loss during probabilistic games (12–15), with an increase in BOLD signal relative to resting baseline for positive outcomes (wins) and a decrease for negative outcomes (losses) (14–16) that resemble learning signals (17, 18). Within the right striatum ROI (20), results differed across win and loss trials. A main effect of incentive ($F_{1,16} = 9.67$, $P < 0.01$) was observed during win trials, driven primarily

by a larger response to monetary reward compared with points reward [$t(16) = 3.11$, $P < 0.01$], but no main effect of social competition was observed. Instead, differences between auction and lottery trials were apparent only in the context of losses ($F_{1,16} = 5.29$, $P < 0.05$). Of particular interest, post hoc t tests showed that mean beta weights for win trials during the auction game (irrespective of incentive) were not significantly different from the lottery game [$t(16) = -0.60$, $P = 0.55$]. In contrast, mean beta weights for losses led to a more pronounced decrease from baseline
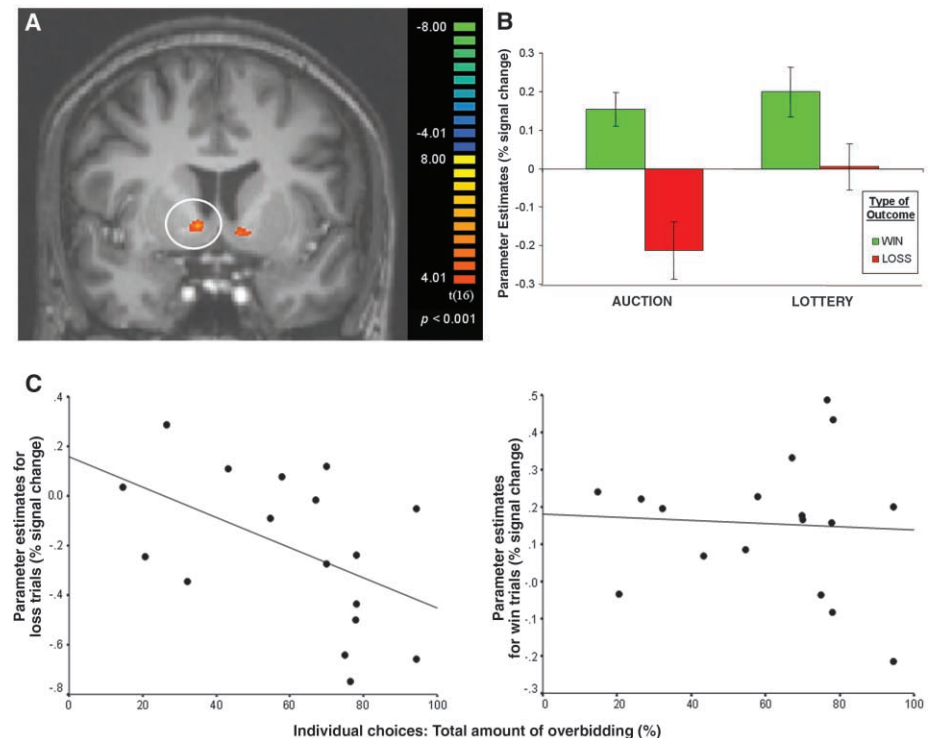
Fig. 1. Striatal response to loss is enhanced by social competition in the auction game. (**A**) Win versus loss outcome contrast: right striatum, including the ventral caudate nucleus was identified as a region of interest (peak at x, y, $z = 10$, 2, 1). (**B**) Parameter estimates, or mean beta weights, for win and loss trials from the right ventral caudate ROI for auction and lottery games show differential responses between auction and lottery only during losses. Error bars are SEM. (**C**) Across participants, the general tendency to overbid during auction trials correlated with BOLD signals in the right striatum ROI (depicted by parameter estimates) when an auction outcome was a loss ($r = -0.493$, $P < 0.05$), but not when the outcome was a win ($r = -0.059$, $P = 0.82$).

Table 1. Results of two separate regressions on the auction data (cluster analysis by subjects) (±SEM). First regression analysis: Overbids are regressed on BOLD signals from the right striatal ROI for individual win-and-loss trials (outcome phase) during the auction game. Second regression analysis: Overbids are regressed on BOLD signals from the rights striatum ROI and on values (V). The number of observations (N) for each trial is given.

Type of trial	N	Regression coefficient	
		Values	BOLD Signal
First regression analysis			
Win trials	397		-0.016 ± 0.1342
Loss trials	372		-0.316 ± 0.2384*
Second regression analysis			
Win trials	397	0.024 ± 0.0208*	-0.04 ± 0.1249
Loss trials	372	0.072 ± 0.0816	-0.237 ± 0.1836*

*$P < 0.05$

during auction compared with lottery trials [$t(16) = -2.30$, $P < 0.05$; (Fig. 1B)]. Finally, a correlation between mean beta weights in the right striatal ROI and a participant's tendency to overbid in general (i.e., the total number of times a participant chose to overbid during auction trials) was observed for loss ($r = -0.493$, $P < 0.05$), but not win ($r = -0.059$, $P = 0.82$) outcomes (Fig. 1C).

Given that there was no actual loss of money or points in either game, it is somewhat surprising that the response to losses during the auction game yielded a significant decrease in BOLD signal relative to the resting baseline and the lottery game. One possibility is that the social competition inherent in the auction game resulted in a loss signal in the right striatum, mirroring that observed with actual monetary loss. The importance of social competition in driving responses in the auction game is further supported by the data from the points condition. Notably, the points incentive could be interpreted as a more relative reward (with respect to other competitors), whereas the monetary incentive is a more abstract reward. A two-way repeated measures ANOVA, including only the data from the points condition with type of social competition and type of feedback (win and loss) as factors, revealed a main effect of competition ($F_{1,16} = 4.44$, $P < 0.05$). This was driven by right striatal responses to winning versus losing during the auc-

tion [$t(16) = 2.01$, $P = 0.06$], rather than the lottery game [$t(16) = 0.50$, $P = 0.63$] (see fig. S1). The finding that social factors can modulate responses in the striatum to monetary incentives has been previously demonstrated with other economic games (21–25). For experimental auctions, it appears that the social interaction specifically alters the response to losses in the striatum, in addition to enhancing overall responses to a nonmonetary reinforcer.

Although the inference of psychological states from BOLD responses should generally be viewed with caution (26), our imaging results provide some initial hypotheses as to the nature of overbidding in experimental auctions. The lack of an enhanced BOLD response in the striatum to wins (in the auction compared with the lottery) suggests that the "joy of winning" may not be mediating overbidding in experimental auctions. In contrast, the stronger BOLD response to losses in the auction game suggests that a fear of losing a social competition may be linked to overbidding. The fear of losing the social competition of an auction may lead to a striatal response similar to that observed in loss aversion (27). However, because no actual losses occurred in this experiment, it would appear that the "fear of losing" the social competition was a factor independent of pure loss aversion.

To further explore these hypotheses a post hoc analysis was conducted. For each subject, we ex-
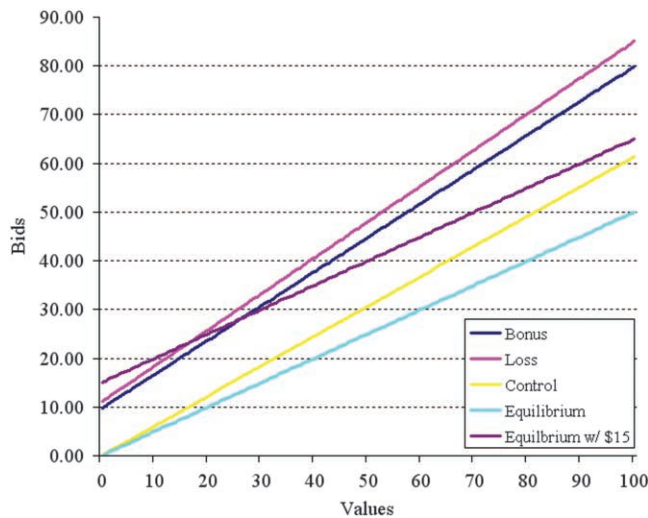
tracted beta weights from the right striatal ROI for wins and losses (outcome phase) for each auction trial. With these beta values, we ran two separate regressions, with a cluster analysis by subjects, examining the relation between the overbids (the difference between the actual bids and the Nash equilibrium), the value assigned (V), and the BOLD response during the outcome phase to a win or loss. Our results indicate that the BOLD response coefficient is not significantly different than 0 for the win trials [$t = -0.25$, $P > 0.05$], but significant responses are observed for loss trials ($t = -2.81$, $P < 0.05$) (Table 1, first regression analysis). The BOLD response regression coefficient of loss trials is significant even when value (V) is included as a controlling factor ($t = -2.74$, $P < 0.05$) (Table 1, second regression analysis). These results, combined with our cross-subject correlation (see Fig. 1C), lead to the intriguing conjecture that perhaps it is the anticipation of a possible loss in experimental auctions that is, at least in part, driving the tendency to bid "too high."

If this conjecture is correct, we should be able to take advantage of this fear of losing to design an experimental auction that will result in an even stronger tendency to overbid. More precisely, if the anticipation of the unpleasant state associated with loss led participants to increase their bids to avoid that state, then manipulating the parameters of a first-price auction to highlight the potential for loss, as opposed to gain, should increase bid values, even if the equilibrium bid function was left unaltered. Hence, we would expect that the loss-frame auction would not only increase bids conditional on value, but also raise more revenue than either a control auction or one where gains or "wins" are emphasized.

In order to test these assumptions, we ran a behavioral economic experiment with three conditions. In all conditions, participants played 30 rounds of an auction game with a randomly assigned single competitor (another participant) on each trial. The range of V was 0 to 100 experimental dollars. In each round, participants knew their own assigned V and the distribution (28) of their competitor's assigned values and were asked to submit a bid (b). The participant submitting the highest b won the good. The payoff was equal to V minus b for the winner and zero for the loser.

There were three experimental groups: baseline, loss-frame, and bonus-frame. The baseline condition was a typical first-price auction as described above. The loss-frame auction was identical to the baseline except that participants were given a sum of 15 experimental dollars at the beginning of each round and were told it was theirs to keep if they won the auction, but that they would have to give it back if they lost. As previously discussed, the purpose of the loss-frame was to prime or enhance the possibility of a loss while hypothesizing, based on the observed striatum BOLD responses, that such priming would increase bidding behavior. The bonus-frame auction was again identical to the baseline, except that participants were told that, in addition to receiving the payoff (V minus b), if they won the auction, they would also be given a bonus

Fig. 2. Estimations of bid functions for the loss-frame, bonus-frame, and baseline control conditions with reference equilibrium functions. Relative to reference equilibrium bids (light blue and purple lines), participants overbid in all three treatment conditions. Consistent with increased overbidding in the loss-frame condition (pink line), bids were higher overall in this condition and the slope of the bidding function was significantly steeper in the loss treatment than in either the bonus-frame (light blue line) ($t = 3.023$, $P < 0.005$) or the baseline control (yellow line) ($t = 11.743$, $P < 0.005$) conditions. In addition, the slope of the bidding function for the bonus-frame condition was significantly steeper than the baseline control condition ($t = 8.283$, $P < 0.005$). See Table 2 for bid functions.

Table 2. Reference equilibrium and estimation of bid strategies for baseline, bonus- and loss-frame conditions. Regressions were conducted with random effects. For estimation of bid strategies, we used a linear specification. Bid functions were estimated using higher-order polynomials, but the coefficients associated with those higher order terms were insignificant.

Condition	Bid function	N	R^2
Equilibrium baseline	$b = 0 + 0.5V$		
Equilibrium + $15	$b = 15 + 0.5V$		
Baseline	$b = 0 + (0.614 \pm 0.011)V$	660	0.805
Bonus-frame	$b = (9.74 \pm 2.547) + (0.702 \pm 0.017)V$	1380	0.733
Loss-frame	$b = (11.09 \pm 2.09) + (0.74 \pm 0.017)V$	1560	0.761

Table 3. Mean revenue by treatment and statistical comparisons (one-tailed *t* test) of revenues across treatments in the behavioral study. Revenue generated in the loss-frame is significantly greater than both the bonus-frame and baseline conditions. In addition, the bonus-frame condition resulted in greater revenue than the baseline condition.

	Revenue	Analysis	
		Baseline	Bonus treatment
Loss treatment	45.62 ± 1.524	$t(1108) = 3.534$ $P = 0.0002$	$t(1468) = 2.943$ $P = 0.0017$
Bonus treatment	42.41 ± 1.495	$t(1018) = -1.26$ $P = 0.1165$	
Baseline	40.88 ± 1.857		

of 15 experimental dollars. Note that in both the loss- and bonus-frame conditions, only the winners get an additional 15 experimental dollars, so the auctions are strategically identical. The difference is simply the way it is framed (*9*). Given this, equilibrium bid functions are the same in the loss-frame and bonus-frame treatments for any given form of the utility function. In the risk-neutral Nash equilibrium for the two nonbaseline treatments, participants' bids should be the same as the baseline condition plus 15 experimental dollars. However, if the hypothesis derived from our fMRI results is correct and the fear of losing is prompting overbidding, we should observe higher overall bids in the loss-frame condition than either the bonus-frame or baseline conditions.

The bid function for each condition is summarized in Table 2. As expected from previous research, there was overbidding in all three conditions relative to the risk-neutral Nash equilibrium. In addition, there was a constant relative increase in overall bid amount in the two nonbaseline conditions due to the additional potential profit of 15 experimental dollars. Consistent with our hypothesis, there was also a significant difference in the slope of the bid functions across conditions. As can be seen in Fig. 2, the bid function for the loss-frame condition is higher overall than the bid function in all other conditions. This is true despite the fact that both the bonus- and loss-frame conditions have identical equilibrium bid functions. If we calculate the actual revenue to a hypothetical auctioneer generated in the experiment, the revenue generated by the loss-frame (45.62) was significantly higher than either the bonus-frame (42.41) or baseline (40.88) conditions (Table 3). By taking advantage of our knowledge of the brain's reward circuitry, we were able to design a novel auction paradigm that led to greater overbidding.

Previous economic investigations of experimental auctions have led to two opposing views as to the nature of overbidding (*5*, *6*). The combination of neuroscience and behavioral techniques provides an interesting perspective on this age-old question. Both our brain imaging and behavioral results are inconsistent with the suggestion that the "joy of winning" mediates overbidding. Although our findings are not inconsistent with a role for risk aversion in the tendency to bid too high, they suggest we should more specifically consider the fear of losing or social loss aversion.

If sensitivity to risk alone is mediating overbidding, then the simple framing manipulation in our behavioral study would not have been effective, because risk was equivalent in both the loss- and bonus-frame conditions. By emphasizing the potential loss in the loss-frame auction, we were able to increase overbidding. Our results suggest that contemplated loss is an important factor in experimental auctions. The fear of losing the social competition inherent in an auction may lead people to pay too high a price for the good for sale. The results of this report, therefore, highlight an extra component in subject's behavior, chiefly the social component of competition, which is not captured by models limited to typical economic variables like profits and probabilities.

Recently, there has been significant debate about whether neuroscience techniques can provide novel insights to economic questions (*1–4*). Although there have been a number of neuroeconomics studies that have utilized economic games to further our understanding of brain function, the benefits to traditional behavioral economics as a result is unclear. As was observed in the progression of cognitive neuroscience, using neuroscience models to inform behavioral or psychological questions requires an initial basic understanding of the neural mechanisms underlying the behavior in question (*3*, *4*, *29*). Because of recent advances in neuroeconomics and our knowledge of the neural circuitry of reward, we were able to leverage our neuroimaging results to develop an auction design that highlights the importance of framing and, specifically, the contemplated loss, as an explanation for overbidding during experimental auctions. Our results provide evidence of how an understanding of the neural systems of economic behavior might inform economic theory.

References and Notes

1. A. Rubinstein, in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, R. Blundell, W. K. Newey, T. Persson, Eds. (Econometric Society 41, Cambridge Univ. Press, Cambridge, vol. 2, 2006), pp. 246–254.
2. F. Gul, W. Pesendorfer, in *Handbook of Economic Methodology*, A. Caplin, A Schotter, Eds. (Oxford Univ. Press, New York, vol. 1, in press).
3. P. W. Glimcher, A. Rustichini, *Science* **306**, 447 (2004).
4. C. F. Camerer, G. Lowenstein, D. Prelec, *J. Econ. Lit.* **43**, 9 (2005).
5. J. C. Cox, B. Roberson, V. Smith, in *Research in Experimental Economics*, vol. 2, V. Smith, Ed. (JAI Press, Greenwich, CT, 1982), pp. 1–43.
6. J. K. Goeree, C. A. Holt, T. R. Palfrey, *J. Econ. Theory* **104**, 247 (2002).
7. R. Cassidy, *Auctions and Auctioneering* (Univ. of Calif. Press, Berkeley, CA, 1967).
8. In auctions, for any given value received, the bidder chooses his optimal bid by trading off the probability of winning (increasing with higher bids) with lower profits if a win occurs. In the trade-off, risk-averse bidders are willing to receive less profit in order to be more certain of a win, hence, overbidding.
9. See supporting online material, available on *Science* Online, for further methodological details including procedures and additional analyses. Each trial was 30 s long, with decisions lasting 4 s, followed by 12 s of fixation and outcomes lasting 2 s, followed by 12 s of intertrial interval.
10. A "loss" refers to losing the bidding phase and not the actual *V* minus *b* payoff. There were a few trials where participants' *b* exceeded the presented *V*, which resulted in an auction "win" but with an associated cost to this excess overbidding (negative final profit). These trials (six overall, across the experiment) were excluded from neuroimaging analysis.
11. See table S3 and supporting online material (SOM) for further analysis. Overbidding was also significant when the incentive was points [$t(16) = 2.40$, $P < 0.05$].
12. B. Knutson, G. W. Fong, S. M. Bennet, C. M. Adams, D. Hommer, *Neuroimage* **18**, 263 (2003).
13. S. Nieuwenhuis *et al.*, *Neuroimage* **25**, 1302 (2005).
14. M. R. Delgado, L. E. Nystrom, C. Fissell, D. C. Noll, J. A. Fiez, *J. Neurophysiol.* **84**, 3072 (2000).
15. M. R. Delgado, H. M. Locke, V. A. Stenger, J. A. Fiez, *Cogn. Affect. Behav. Neurosci.* **3**, 27 (2003).
16. E. Tricomi, M. R. Delgado, B. D. McCandliss, J. L. McClelland, J. A. Fiez, *J. Cognit. Neurosci.* **18**, 1029 (2006).
17. M. Haruno *et al.*, *J. Neurosci.* **24**, 1660 (2004).
18. J. P. O'Doherty, *Curr. Opin. Neurobiol.* **14**, 769 (2004).
19. Mean beta weights are parameter estimates derived from the general linear model that indicate how much each factor (i.e., individual condition) contributes to the overall data. For detailed explanation see (*30*).
20. Similar results for the left ventral caudate ROI are reported in the SOM.
21. D. J. de Quervain, *Science* **305**, 1254 (2004).
22. E. Fehr, B. Rockenbach, *Curr. Opin. Neurobiol.* **14**, 784 (2004).
23. B. King-Casas *et al.*, *Science* **308**, 78 (2005).
24. P. R. Montague, B. King-Casas, J. D. Cohen, *Annu. Rev. Neurosci.* **29**, 417 (2006).
25. M. R. Delgado, R. H. Frank, E. A. Phelps, *Nat. Neurosci.* **8**, 1611 (2005).
26. R. A. Poldrack, *Trends Cognit. Sci.* **10**, 59 (2006).
27. S. M. Tom, C. R. Fox, C. Trepel, R. A. Poldrack, *Science* **315**, 515 (2007).
28. The distribution of values for all bidders was uniform over the four values available (i.e., each value had an equal probability of being chosen).
29. E. A. Phelps, in *Nature of Cognition*, R. Sternberg, Ed. (MIT Press, Cambridge, MA, 1999), pp. 295–313.
30. S. A. Huettel, A. W. Song, G. McCarthy, *Functional Magnetic Resonance Imaging* (Sinauer Associates, Sunderland, MA, 2004), p. xviii.
31. The authors would like to acknowledge support from the James S. McDonnell Foundation and National Institute of Mental Health, NIH (MH62104) to E.A.P. and the support of the Seaver Foundation to NYU's Center for Brain Imaging. We also thank the Center for Experimental Social Science for its support. We are grateful to D. Schwarz for his efforts in the initial stages of collection and analysis of neuroimaging data. We thank J. Zakrzewski and K. Nearing for their assistance during data collection, M. Niznikiewicz for assistance during data analysis, and R. Advani for his programming help on the behavioral experiment.

Stirring System

Viscous reaction mixtures often require more powerful stirring than can be achieved with magnetic stirrers. The Vortex stirrer system converts a single overhead stirrer into a powerful three-position parallel stirrer combined with a multiple heating block. The combination of Vortex stirring and DrySyn heating blocks provides efficient stirring while retaining the benefits of the DrySyn range, including safety, rapid heat transfer, and a small size. Used with a standard laboratory hotplate, the unit becomes a compact, small-footprint reaction station capable of running up to three parallel reactions in conventional round-bottom flasks.

Asynt
For information +44-(0)-1638-781709
www.asynt.com

Data Management for Sequencing Platforms

The HT-BLIS productivity system provides an enterprise-scale informatics environment for integrating, analyzing, and sharing data generated by next-generation sequencing systems. HT-BLIS can process, in real time, the massive amounts of data generated by high throughput genomic analysis systems such as Applied Biosystems' Solid System, Helicos's HeliScope Sequencer, Illumina's Genome Analyzer, and Roche's Genome Sequencer FLX System. The latest genomic analysis technologies generate millions of sequences with billions of bases per day. While this capability offers unprecedented resolution and speed for gene expression, sequencing, and genotyping projects, the huge data volumes overwhelm existing informatics infrastructures. HT-BLIS addresses this challenge by receiving, storing, and manipulating in real time the massive amounts of data generated by banks of next-generation sequencers.
Biotique Systems
For information 775-787-9900
www.biotiquesystems.com/HT-BLIS

Handheld X-Ray Fluorescence Instrument

The TracerturboSD is a handheld X-ray fluorescence instrument that features a silicon drift detector (SDD) for dramatically improved speed, sensitivity, and resolution. Bruker's XFlash SDD has been integrated into the handheld device to achieve the speed, analytical specificity, and energy resolution that is generally found only in more expensive laboratory systems. New applications include art conservation, archeology analysis, and the aerospace industry, where users can better analyze light element alloys, even without the use of vacuum or helium attachments.
Bruker AXS
For information 609-847-9468
www.bruker-axs.com

End Repair Kit

The End-It DNA End Repair Kit is for treating fragmented genomic DNA for use in next-generation DNA sequencing. The kit converts DNA containing damaged ends, 5' protruding ends, or 3' protruding ends to 5'-phosphorylated, blunt-ended DNA for fast and efficient ligation to sequencing adapters. The end-repaired, 5'-phosphorylated DNA can then be efficiently ligated using Epicentre's Fast-Link DNA Ligation Kit.
Epicentre Biotechnologies
For information 800-284-8474
www.EpiBio.com

Electrochemical Detection Guide

ESA Electrochemical Detection Guide is a 14-page brochure that leads the reader through the features, benefits, and technical specifications of high-performance electrochemical detectors. It describes systems optimized for applications ranging from identifying oxidation or metabolic products of small molecules to making fast measurements of neurotransmitters in discrete brain regions. These systems offer a wide range of detection options, advanced amperometric and coulometric cells, pumps that minimize background noise, data stations, and high-precision autosamplers.
ESA Biosciences
For information +44-1844-239381
www.esainc.com

Freeze Drying System

The Virtis Genesis Freeze Drying System offers the versatility for use in pilot, research, and small-scale applications. The compact, freestanding, mobile design enables the system to be optimally configured to almost any freeze drying application. New ergonomic profiling allows easy inspection of the vacuum pump and facilitates quick, trouble-free oil changes. A clean room version is also available. The product chamber, shelves, and condenser chamber are made of durable stainless steel, with a squared product chamber that ensures easy cleaning and maximum shelf area. A four-inch diameter port increases vapor flow from the product to the condenser chamber, maximizing productivity. It is available with up to six bulk-drying shelves and five vial processing shelves.
Genevac
For information +44-1473-240000
www.genevac.com